Univ. of
Toronto
Library

BINDING LIST JUL 1 1922

Digitized by the Internet Archive
in 2009 with funding from
University of Toronto

http://www.archive.org/details/matematiskfysisk02dans

MATHEMATISK-FYSISKE

MEDDELELSER

UDGIVNE AF

DET KGL. DANSKE VIDENSKABERNES SELSKAB

2. BIND

KØBENHAVN

HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL

BIANCO LUNOS BOGTRYKKERI

1919—20

Q
62
D28
bd. 2

INDHOLD

1. The photochemical Decomposition of Hydrogen Peroxide. By Chr. Winther.
2. The photochemical Oxidation of Hydriodic Acid. By Chr. Winther.
3. The photochemical Efficiency of the absorbed Radiation. By Chr. Winther.
4. Sur l'origine de l'algèbre. Par H. G. Zeuthen.
5. Talet. Inledning till teorien för analytiska funktioner. Af G. Mittag-Leffler.
6. Experimentalundersøgelser over Gnidningselektricitetens Oprindelse. VII. Af C. Christiansen og Johanne Christiansen.
7. Experimentalundersøgelser over Gnidningselektricitetens Oprindelse. VIII. Af C. Christiansen.
8. Overfladespændingens Indflydelse ved Udstrømning af en Vædske i Straaleform. Af Jul. Hartmann.
9. Über den Nachweis von Chlorid neben Bromid. Von Carl Faurholt.
10. On the Solubility of Salts in Salt Solutions. Studies on Solubility I. By J. N. Brønsted. With 10 figures in the text.
11. Die kausale Relativitätsforderung und Einsteins Relativitätstheorie. Von Helge Holst.
12. Recherches sur les Polynomes de Stirling. Par Niels Nielsen.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser II, 1.

THE PHOTOCHEMICAL DECOMPOSITION OF HYDROGEN PEROXIDE

BY

CHR. WINTHER

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL.
BIANCO LUNOS BOGTRYKKERI
1919

Introduction.

In 1900 KISTIAKOWSKY[1] found that potassiumferro and ferri cyanide were able to act as an optical sensitiser on the decomposition of hydrogen peroxide. This process exhibits several interesting features. It shows, f. inst., a very marked after-action. KISTIAKOWSKY's measurements showed that a quite evanescent insolation (1 minute) was sufficient to give the same velocity as a continued insolation, a fact which suggests the formation by the insolation of a catalyst of very great longevity. The catalyst formation was directly shown by WEIGERT[2], by first exposing a solution of potassiumferro-cyanide to light and then adding it to hydrogen peroxide in the darkness, whereafter a vivid decomposition followed. J. AMANN[3], too, was able to show ultra-microscopically that by exposure of potassiumferro-cyanide, colloid particles are formed, by which the oxygen-formation of the hydrogen peroxide mainly takes place. The mechanism of the process, hence, seems fairly well elucidated, if we disregard the fact that no determinations of colour-sensitiveness are present.

With regard to the order of reaction, KISTIAKOWSKY found that the process in special cases, i. e. at certain mixtures of ferro- and ferri-cyanide, was mono-molecular, but that the constant in all other cases gradually increased somewhat. This

[1] Zeitschr. f. physikal. Chemie **35**, 431 (1900).
[2] Ann. d. Physik (4) **24**, 261 (1907).
[3] Kolloid-Zeitschr. **8**, 13 (1911).

vagueness cannot give rise to any astonisment, because of the exclusive employment of mixed light (arc) for the exposure.

By trying to prepare this process for lecture purposes it proved, however, to my great astonishment, that the reaction velocity was constant everywhere, the process, thence, was zero-molecular. At my desire Mr. S. Weibel therefore examined a series of different concentrations of hydrogen peroxide, potassiumferro- and potassiumferri-cyanide, amongst which also those for which Kistiakowsky had found the first order. But the velocity was found to be constant everywhere both in monocromatic and mixed light, with arc and with quartz-mercury-lamp, and under extremely different conditions with regard to the exposure of the single portions and the taking out of samples, and we have not succeeded in finding the cause of the discrepancies between Kistiakowsky's and our results.

The final measurements which were carried out by myself, were for the special purpose of 1) determining the efficiency of the absorbed radiation, 2) to find by which law the light-catalyst is formed and disappears and 3) to calculate the rate of reaction in the mixture of hydrogen peroxide and potassiumferro-cyanide from the knowledge of the formation and effect of the catalyst comprised by the two previous points.

The employed arrangement was quite the same as by the quantitative investigation of the oxidation of the hydriotic acid[1], only differing by the omission of the air-supply which was compensated by mechanical stirring by means of a very rapidly rotating spiral glass stirrer. All these exposures were executed in a couvette of 1 cm. thickness of layer and furnished with glass walls. The quartz-mercury-lamp, with in-

[1] Kgl. Danske Vid. Selsk, Math.-fys. Medd. II, 2.

serted ray filters, served as a light-source. In order to obtain constant results it proved necessary to cleanse all the glass apparatus with strong hydrochloric acid.

Formation of the Catalyst.

For this 0,075 gm. potassiumferro-cyanide was dissolved in 50 cm.^{3} water (0,0035-m). After having exposed this solution for t minutes, 5 cm.^{3} was taken out and added to 5 cm.^{3} diluted hydrogen peroxide (abt. 0,08-m and 0,04-m). The mixture was left in darkness for 30 minutes and was then acidified with sulphuric acid and titrated with 0,02-m potassium permanganate.

Table 1.

405 μμ (Fig. 1).

Experiment 1. 360 Watt, $E = 4{,}7$ Erg.			Experiment 2. 340 Watt, $E = 4{,}4$ Erg.		
t	$KMnO_4$	$\Delta H_2O_2.10^4$	t	$KMnO_4$	$\Delta H_2O_2.10^4$
0	20,10	—	0	20,25	—
3	20,00	0,02	3	20,15	0,02
6	19,80	0,06	6	20,05	0,04
9	19,60	0,10	9	19,85	0,08
12	19,60	0,10	12	19,85	0,08
15	19,10	0,20	15	19,35	0,18
18	19,35	0,15	18	19,65	0,12

366 μμ (Fig. 2 and 3).

Experiment 3. 330 Watt, $E = 6{,}4$ Erg.			Experiment 4. 410 Watt, $E = 6{,}6$ Erg.			Experiment 5. 440 Watt, $E = 7{,}0$ Erg.		
t	$KMnO_4$	$\Delta H_2O_2.10^4$	t	$KMnO_4$	$\Delta H_2O_2.10^4$	t	$KMnO_4$	$\Delta H_2O_2.10^4$
0	20,20	—	0	20,22	—	0	20,30	—
3	19,00	0,24	3	18,55	0,33	3	18,10	0,44
6	18,60	0,32	6	18,25	0,39	6	17,80	0,50
9	18,60	0,32	9	18,45	0,35	9	17,90	0,48
12	18,35	0,37	12	18,35	0,37	12	18,00	0,46
15	18,45	0,35	15	18,45	0,35	15	18,18	0,42
18	18,40	0,36	18	18,30	0,38	18	18,02	0,46

Experiment 6. 415 Watt, $E = 13{,}9$ Erg.				Experiment 7. 390 Watt, $E = 11{,}2$ Erg.			Experiment 8. 430 Watt, $E = 12{,}9$ Erg.		
t	$KMnO_4$	$\Delta H_2O_2.10^4$	k	t	$KMnO_4$	$\Delta H_2O_2.10^4$	t	$KMnO_4$	$\Delta H_2O_2.10^4$
0	19,75	—	—	0	10,10	—	0	10,45	—
1	18,60	0,23	0,18	1	9,45	0,13	1	9,77	0,14
2	17,70	0,41	0,21	2	8,80	0,26	2	8,90	0,31
3	17,05	0,54	0,25	3	8,55	0,31	3	8,60	0,37
4	16,70	0,61	0,25	4	8,20	0,38	4	8,50	0,39
6	16,63	0,62	0,22	6	8,30	0,36	6	8,35	0,43
8	16,60	0,63	0,19	8	8,15	0,39	8	8,45	0,40
	$c_\infty = 0{,}65$		0,22						

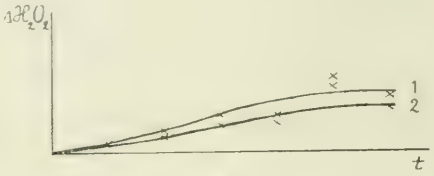


Fig. 1.

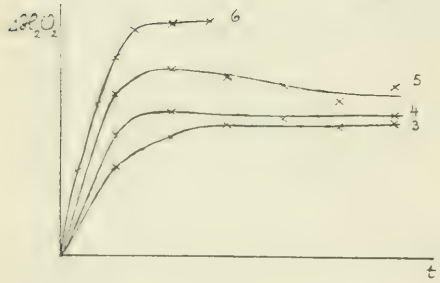


Fig. 2.

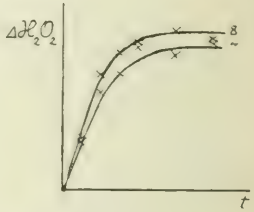


Fig. 3.

The experiments show that the catalyst formation increases with the time of exposure until a certain maximum concentration, which rises with the applied energy of light, and — to a far slighter degree — with the hydrogen peroxide-concentration. This maximum concentration for the hydrogen peroxide-concentration 0,08-m (Exp. 1—6) can with good approximation be put equal to:

$$c_\infty = 0{,}071 \cdot 10^{-4} \cdot E_{abs} + 0{,}14 \cdot 10^{-4}$$

where E_{abs} is the energy absorbed per mm² and second in erg. The agreement is seen in table 2.

Table 2.

Experiment	λ	E_{abs}	α found	α calcul.	α diff.
1	405	0,31	0,15	0,16	—0,01
2	»	0,29	0,12	0,16	—0,04
3	366	3,3	0,36	0,37	—0,01
4	»	3,4	0,37	0,38	—0,01
5	»	3,6	0,46	0,40	+0,06
6	»	7,1	0,65	0,64	+0,01

Also the velocity with which the catalyst is formed increases, as shown by the figures, with the intensity, and the initial velocity seems to be proportional to the absorbed energy of light.

The fraction of the penetrating light, absorbed in 1 cm. layer was found to be 0,065 for λ405 (with spectral photometer) and for λ366 (with fluormeter) 0,51. In the experiments 1—6, the amount of substance transformed in the beginning per absorbed energy-unit (erg.) will in mol. 10^{10} hereafter be:

Table 3.

Experiment	λ	t	$\Delta H_2O_2 \cdot 10^4$	E_{abs}	$(\Delta H_2O_2 : E_{abs}) \cdot 10^{10}$
1	405	3	0,02	$2{,}1 \cdot 10^4$	1,0
2	»	3	0,02	2,1 »	1,0
3	366	3	0,24	$2{,}9 \cdot 10^5$	0,8
4	»	3	0,33	3,0 »	1,1
5	»	3	0,44	3,2 »	1,4
6	»	1	0,23	2,1 »	1,1
					1,1

The experiments 7 and 6 (half hydrogen peroxide-concentration) give the following values:

7	366	1	0,13	$1{,}7 \cdot 10^5$	0,8
8	»	1	0,14	1.9 »	0.7
					0,75

The efficiency is here, perhaps, a little lower than for the double concentration. For the most complete and accurate of the velocity curves (experiment 6) a fairly good reaction

constant of the first order may be calculated (as seen in the table). The velocity formula may therefore be written:

$$\frac{dc}{dt} = k_1 E_{abs} - k_2 c$$

where E_{abs} is the absorbed amount of energy, c the concentration of the catalyst to the time t, k_1 and k_2 the velocity constants by the formation and transformation of the catalyst. The found constant — typified in the table — is then = 0.4343 k_2 and refers to the retrogressive transformation of the catalyst.

The Transformation of the Catalyst.

The magnitude 0,4343 k_2 in mean being equal to 0,22, it is easy to calculate that the retrogressive transformation in darkness ought to occur so rapidly, that after the course of 1 minute 60 per cent., after 3 minutes 22 per cent., and after 10 minutes 0,6 per cent. should be left of the maximum amount of catalyst. These predictions prove very far from the mark here.

The 0,0035-molar potassiumferro-cyanide-solution was insolated for 8 minutes with the wave length 366 μμ (hence, to maximum catalyst-concentration). After t minuts, 5 cm.^{3} were taken out in darkness and added to 5 cm.^{3} diluted hydrogen peroxide (abt. 0,08-m). After having been left for 30 minutes, the mixture was acidified and titrated.

Table 4.

Experiment 9. 420 Watt.		Experiment 10. 430 Watt.		Experiment 11. 430 Watt.	
t	$\Delta H_2O_2 \cdot 10^4$	t	$\Delta H_2O_2 \cdot 10^4$	t	$\Delta H_2O_2 \cdot 10^4$
0	0,36	0	0,56	0	0,30
1	0,47	10	0,34	3	0,25
2	0,49	20	0,30	6	0,31
3	0,47	30	0,38	9	0,33
4	0,45	40	0,28	12	0,26
6	0,47	60	0,27	15	0,28

It is only in the beginning of experiment 10 that a decrease in the amount of catalyst seems to take place, but thereafter this quantity keeps fairly constant instead of falling to zero. In the two other experiments the amount of catalyst oscillates irregularly around a mean value. The first value in experiment 10 must then, presumably, be due to an experimental error, inasmuch as the subsequent section shows that the catalyst does not decrease perceptly in the course of several hours.

The apparent disagreement between this extraordinary stability and the existence of a stationary state might, perhaps, be explained in the following way. KISTIAKOWSKY has found that a potassiumferro-cyanide-solution becomes alkaline by insolation. But according to HENRI and WURMSER[1] the decomposition of hydrogen peroxide in light is inhibited by bases, whereas the decomposition of the same substance in darkness is accelerated by bases. Insolated potassiumferro-cyanide-solution must with its contents of light catalyst be considered as light in regard to hydrogen peroxide, and there is consequently the possibility that the alkali formed by the insolation can act as a negative catalyst to such a degree, that a compromise occurs between the effect of the catalyst and the hydroxylion, the concentrations of which both increase by continued exposure.

It is difficult to know what quantities can come into question. Taking for granted what is certainly exaggerated, that each molecule of potassium ferro-cyanide transformed to light catalyst thereby yields a molecule of potassium hydroxide, we can by means of the law of critical efficiency mentioned later on[2], calculate that, f. inst. in experiment 6 0.00035 mol. potassium hydroxide per liter will be formed in the

[1] Compt. rend. 157,284 (1913).

[2] Kgl. Danske Vid. Selsk. Math.-fys. Medd. II, 3.

course of 4 minutes. According to the results of Henri and Wurmser such an alkali concentration will lower the decomposition of the hydrogen peroxide to abt. 30 per cent. of the value in neutral solution, it would, then, act as a powerful negative catalyst.

This explanation cannot after all be correct, for as the inhibiting effects of the potassium hydroxide increase far more than its concentration, while the light catalyst must be supposed to be formed with constant velocity, the absorption remaining practically constant, the effects on the hydrogen peroxide should necessarily first pass a maximum and then fall in a marked degree by further exposure, while the experiments point decisively in the direction of a stationary state.

Another explanation might be that an inhibiting substance was formed simultaneously with and proportionally to the catalyst itself. If the inhibiting effect of this substance be proportional to its concentration, we obtain a velocity formula of similar form as the one discussed above, with the difference only that the constant k_2 now — one constant factor excepted — corresponds to the formation of the catalyst.

Above was found:

$$c_\infty = 0{,}071 \cdot 10^{-4} E_{abs} + 0{,}14 \cdot 10^{-4}$$

The velocity formula which in the stationary state should give this result, must be:

$$\frac{dc}{dt} = K\,(0{,}071 \cdot 10^{-4}\ E_{abs} + 0{,}14 \cdot 10^{-4} - c).$$

From experiment 6 we have:

$$K = \frac{0{,}23 \cdot 10^{-4}}{0{,}071 \cdot 10^{-4} \cdot 7{,}1 + 0{,}14 \cdot 10^{-4}} = 0{,}35.$$

By application of this formula we obtain the following values for the quantity transformed in the first minute:

Experiment:	1	2	3	4	5
	0,05	0,05	0,13	0,13	$0{,}14 \cdot 10^{-4}$.

while in 3 minutes was found:

0,02 0,02 0,24 0.33 $0.44 \cdot 10^{-4}$

Table 5 shows the accordance between the found and the calculated values for experiment 6.

Table 5.

$H_2O_2 \cdot 10^4$	
found	cal.
0,23	(0,23)
0,41	0,38
0,54	0,48
0,61	0,54
0,62	0,61
0,63	0,63

The formula is evidently fairly well adapted for the experiments with rather large decompositions, whereas it quite fails us for the small alterations in experiments 1 and 2, where they are found as very small differences between two titrations having fairly large quantities of liquid, and they cannot be very correct, which is also shown by the two said experimental tables. Any importance can, therefore, hardly be attached to this disagreement.

Considering the formula to be correct, we obtain for the initial velocity no simple proportionality with the absorbed energy, as the experiments seemed to show, but a lineary relation expressed by:

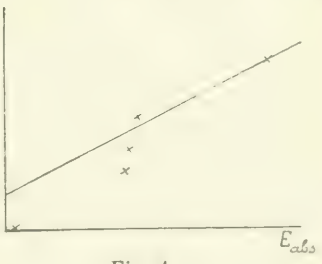


Fig. 4.

$$\frac{dc}{dt} = 0{,}35\ (0{,}71 \cdot 10^{-4}\ E_{\text{abs}} + 0.14 \cdot 10^{-4})$$

Fig. 4 contains this line together with the found values calculated for the first minute, which no doubt gives too small values.

What significance the constant magnitude $0.14 \cdot 10^{-4}$ possesses is an unsolved question. The mixture of hydrogen peroxide and potassiumferro-cyanide showing, also in darkness, a faint decomposition, under the given circumstances being

abt. $0{,}13 \cdot 10^{-4}$ in 30 minutes. it is an obvious conclusion, that the potassiumferro-cyanide-solution already previously to the exposure contains a corresponding amount of the colloidal catalyst. But this darkness-process is beforehand subtracted from all the given numbers, and it is not easy to see how this originally present catalyst could act twice as powerfully by exposure as in darkness, all the experiments showing that the catalyst formed by exposure acts with exactly the same power in darkness as in light, as soon as it has been formed. The assumption that the nuclei of the darkness catalyst should be able to serve as specially fruitful starting points for the formation of light catalyst, only introduces the absorbed energy also in this term of the formula, deprives it, then, of its constancy, and is consequently incapable of removing the difficulty.

The Action of the Catalyst.

In all of the previously adduced experiments the decomposition of the hydrogen peroxide was measured after the mixture with potassiumferro-cyanide had been left in darkness for 30 minutes. This period was chosen for purely practical reasons and is, as will be shown by the following experiments, as applicable as any other period within reasonable limits.

Potassiumferro-cyanide solution (0,0035-m) was in 1 cm. layer exposed to light for 3 minutes with $\lambda 366$, and thereafter added to equal volumes of hydrogen peroxide-solution (0,08-m). The mixture was left in darkness for t minutes, after which 10 cm³ were taken out and titrated.

Table 6.
Experiment 12.
400 Watt, $E = 13{,}2$ Erg.

t	5	10	20	40	60	
$\Delta H_2O_2 10^4$	0.12	0.20	0.37	0.61	0.90	
$\frac{\Delta H_2O_2 10^4}{t}$	0.024	0.020	0.019	0.015	0.015	Cal. 0.021

In another experiment with 2 minutes primary exposure was found:

Experiment 13.

416 Watt, $E = 14{,}2$ Erg.

t	30	60	120	150	
$\Delta H_2O_2 10^4$	0,40	1.23	2.67	3,04	
$\frac{\Delta H_2O_2 10^4}{t}$	0,013	0,021	0,022	0,020	Cal. 0,022

A portion, being quite uninsolated, in 150 minutes lost $0{,}13 \cdot 10^{-4}$ mol H_2O_2.

A third experiment with 8 minutes primary exposure gave:

Experiment 14.

445 Watt, $E = 19{,}5$ Erg.

t	30	60	120	180	
$\Delta H_2O_2 10^4$	0,44	0,73	1,83	3,04	
$\frac{\Delta H_2O_2 10^4}{t}$	0,015	0,012	0,015	0,017	Cal. 0,029

These results are exhibited in fig. 5 from which it appears that the catalyst acts with almost constant velocity until the greater portion of the hydrogen peroxide is decomposed. The decrease of velocity which then appears may very well be due to the mass action, so that the catalyst, then, does not presumably loose any of its activity by the decomposition itself, but keeps this activity unaltered through several hours and perhaps through a much longer period. The values for the decomposition per minute which are given under »cal.» are calculated from the formula for c_∞ (v. p. 6), which implies that the formation of catalyst has been complete.

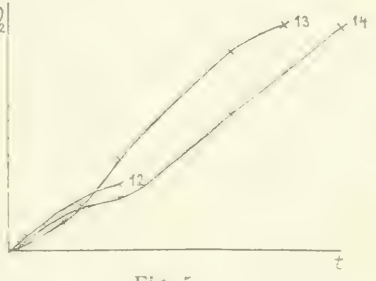


Fig. 5.

The Total Process.

When the velocity with which the catalyst is formed by exposure as well as the constant velocity with which the catalyst exerts its action are both known, the velocity of the total process should be computable from this, if new complications do not appear by exposure of the mixture of the two substances. The provisional experiments indeed pointed in the direction of the presence of such complications, the mixture rather quickly becoming much darker (greenish) coloured by exposure, whereas the pure potassiumferro-cyanide-solution keeps its colour unaltered. An experiment to determine the extinction-coefficient of the mixture for $\lambda 405$ for this reason remained without result. During the measurement itself with the spectral photometer the magnitude αc increased from 0,55 to 0,79 in the course of abt. 10 minutes, whereby the quartz-mercury lamp, screened with dull glass, served as a light-source. For the corresponding pure solution of potassiumferro-cyanide is $\alpha c = 0{,}029$, and it has been impossible for me to decide how much of the increase up to 0,55 is due to the mixing process itself and how much to the exposure.

In spite of this considerable increase in the absorption, the mixture is only very little sensitive to $\lambda 405$. The found losses of hydrogen peroxide are so small and vary so irregularly that there is no reason to state the numbers.

With regard to $\lambda 366$ it is quite another matter. The estinction was measured behind the spectograph by means of the fluormeter. There was measured partly a fresh uninsolated mixture, and partly a mixture having served for one of the experimental exposures given below, whereby abt. half of the hydrogen peroxide was decomposed. In the first case (fresh

solution) $\alpha c = 0{,}31$, in the last case (exposed solution) 0,43 while the pure solution of potassiumferro-cyanide has $\alpha c = 0{,}31$. In this case there is thus only a comparatively slight increase of the absorption by exposure, and the great increase of $\lambda 405$ must therefore be due to quite a new absorption band in this part of the spectre, a band which has evidently nothing to do with the formation of the light catalyst. We might therefore hope to avoid complications by only exposing the mixture to $\lambda 366$. Two such experiments of exposure were carried out. In both cases 50 cm.^{3} potassiumferro-cyanide-solution (0,0035-m) was mixed with 50 cm.^{3} hydrogen peroxide (0,08-m) and exposed for t minutes in 2 cm. layer. Thereafter 10 cm.^{3} were taken out and titrated immediately.

Table 7.

Experiment 15.
440 Watt, $E = 15{,}9$ Erg.

t	$\Delta H_2O_2 10^4$ found	calculated I	II	III
10	0,15	0,19	0,25	0,20
20	0,53	0,44	0,58	0,60
43	1,39	1,02	1,34	1,28
68	2,09	1,65	2,17	2,10

Experiment 16.
405 Watt, $E = 16{,}6$ Erg.

t	$\Delta H_2O_2 10^4$ found	calculated I	II	III
5	0,04	0,08	0,10	0,08
10	0,18	0,19	0,25	0,20
15	0,36	0,32	0,42	0,33
25	0,74	0,57	0,75	0,74
35	1,19	0,82	1,08	1,04

As shown particularly by experiment 16 and fig. 6, there is also in this process — as in the oxidation of iodine hydrogen[1] — a period of induction arising here as there, from the fact that the catalyst is formed slowly by the exposure itself. There is, however, the difference between the two processes, that in the case before us, the substance concerned — potassium-

Fig. 6.

[1]) Kgl. Danske Vid. Selsk., Math.-fys. Medd. II, 2.

ferrocyanide — is light-sensitive. the catalyst thus being able to be formed even if none of it is found beforehand, whereas it is an absolute condition of the sensitivity of the hydriotic acid that some of the catalyst is found beforehand, the hydrogen iodide not being sensitive in itself.

The magnitudes given in the columns »calculated, I, II and III« are calculated in the following manner.

The velocity of reaction by the formation of the catalyst was found above (p. 10) to be:

$$\frac{dc}{dt} = 0{,}35\ (0{,}071 \cdot 10^{-4}\ E_{abs} + 0{,}14 \cdot 10^{-4} - c)$$

from which:

$$c = (0{,}071 \cdot 10^{-4}\ E_{abs} + 0{,}14 \cdot 10^{-4})\ (1 - 10^{-0{,}4343 \cdot 0{,}35\, t}).$$

In the way this formula is derived the velocity is expressed by the amount of hydrogen peroxide which is decomposed by being left for 30 minutes with the newly formed catalyst. This velocity of decomposition for the hydrogen peroxide is — as shown above — constant. In 1 minute, consequently, $^1/_{30}$ of the amount which is denoted by the above formula is decomposed. If no unforeseen complications arise by exposure of the mixed solution, the hydrogen peroxide must by this exposure be decomposed with the same velocity as by primary exposure of potassiumferro-cyanide, and subsequent mixing with hydrogen peroxide in darkness, the activity of the catalyst being apparently of unlimited duration. The velocity of decomposition of the hydrogen peroxide by exposure of the mixed solution is consequently:

$$-\frac{dc'}{dt} = \frac{1}{30}\, c = (0{,}0024 \cdot 10^{-4}\ E_{abs} + 0{,}0047)\ (1 - 10^{-0{,}4343 \cdot 0{,}35\, t}).$$

Column I in the above table is calculated on the presupposition that the estinction of the mixture has kept unaltered ($\alpha c = 0{,}31$) during the exposure; the calculated figures being so much more behind the found the longer the exposure lasts. It is taken for granted in the calculation of column II that the estinction has all the time had the slightly greater value which was found for the exposed mixture ($\alpha c = 0{,}43$). In this place the calculated figures are too large to start with, but agree so much the better with the found ones, the longer the exposure lasts. Assuming, finally, the alteration from 0,31 to 0,43 to have taken place proportionally to the time in the 68 minutes the mixture has been exposed before the estinction was measured last time, we obtain column III, which agrees as well as can be expected with the found values.

The velocity of reaction by exposure of the mixture of potassiumferro-cyanide and hydrogen peroxide can consequently be calculated when the velocity by which the light-catalyst is formed, and the velocity with which this light-catalyst decomposes hydrogen peroxide in darkness are known.

Results of This Work.

By exposure of a mixture of potassium ferrocyanide and hydrogen peroxide in diluted solution with ultra-violet the hydrogen peroxide is decomposed at a rate, which at first is very small, afterwards increases, and finally becomes constant.

This is explained by the slow formation of a constant amount of a substance, which catalyses the decomposition of hydrogen peroxide, and the stability of which is very great.

The rate of decomposition of hydrogen peroxide in the

mixed solution can be calculated, when the rate of formation of the catalyst and the rate of its effect on the hydrogen peroxide is known.

Royal Technical College of Copenhagen,
May 1919.

Færdig fra Trykkeriet den 10. Februar 1920.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. II, 2.

THE PHOTOCHEMICAL OXIDATION OF HYDRIODIC ACID

BY

CHR. WINTHER

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SON, KGL. HOF-BOGHANDEL.
BIANCO LUNOS BOGTRYKKERI
1920

I. The photochemical Oxidation of the hydriodic acid.

Introduction.

The earlier literature on this photolysis is found embodied in EDER's Photochemistry[1]. Of more recent works PLOTNIKOW's two papers[2] have more especially given the impression of enlarging our knowledge of the said process to a considerable degree. PLOTNIKOW's experiments were carried out by exposure to an uviol-lamp of potassium iodide-solutions, to which sulphuric acid and water containing oxygen had been added. What was being measured was, then, the decrease in the oxygen-concentration. No stirring was employed. Neither induction-period nor after-action was found. The velocity constant is proportional to the light-intensity. The rates of the dark- and light processes are added directly and occur after the same order of reaction, so that the velocity of the total process may be expressed by:

$$-\frac{d(O_2)}{dt} = [K_d 10^{\alpha T} + K_L 10^{\beta T}]\,(HCl)^{2/3}(KJ)^{2/3}(O_2).$$

K_d and K_L are the velocity constants for the dark- and light-processes, $10^{\alpha T}$ and $10^{\beta T}$ the corresponding temperature-coefficients.

PLOTNIKOW furthermore found that the dark- and light-process behaved quite differently in regard to catalysts, the selfsame substance, ($CuSO_4$ f. inst.), being able to accelerate

[1] 3d Edition. P. 144 (1906).

[2] Zeitschr. f. physichal. Chemie. **58**, 214 (1907), **64**, 215 (1908).

the one and inhibit the other; that the process is especially sensitive to blue light, less to violet and ultra-violet (336 μμ); that an addition of iodine to the reaction-mixture does not abate the velocity; and finally, that the reaction-mixture itself, employed as a light-filter, does not reduce the velocity of another portion of reaction-mixture which is placed behind this light-filter. From this Plotnikow concludes, that a very slightly absorbing reaction-mixture is here in question, which despite this has a considerable sensitiveness. This interpretation, which is generally accepted, is expressed by Bodenstein in his survey[1] on quantum-sensitivities, where he computes the amount of molecules oxidized in aqueous hydriodic acid for each absorbed quantum to 10^6, the same order of magnitude, hence, as the sensitiveness of the mixture of chlorine and hydrogen.

These statements of Plotnikow's contain, however, such extraordinary and improbable things, that it seems highly necessary to submit his work to a critical examination.

In the first place, the found colour-sensitiveness seems to be but little probable. In the course of time not a few cases are known where the colour-sensitiveness occurs parallel to (and is often proportional to) the absorption of the light-sensitive substance, and several of the apparent exceptions from this quantitative absorption-law can be explained from the particular circumstance sprevailing in the individual case[2]. But in this reaction-mixture no single substance is found, the absorption of which is greater for blue than for violet and nearest ultra-violet. Neither was the confidence in Plotnikow's results increased on a previous occasion[3]

[1] Zeitschr. f. physichal. Chemie **85**, 356 (1913).
[2] V. f. inst. Lasareff. Ann. d. Physik (4) **37**, (1912) and below S.
[3] Zeitschr. f. Elektroch. **19**, 394 (1913).

when I found that the light-filters employed by him were far from being monocromatic.

Then: If really, as generally assumed, hydriodic acid, or rather iodine-ion, is the actual sensitive substance, then the new formed iodine, or rather the tri-iodine-ion, the absorption of which for blue, violet and nearest ultra-violet is infinitely stronger than that of the iodine-ion, must absolutely reduce the light-intensity in the interior of the liquid to such a degree, that the process for this reason must come to a stand-still long before the oxygen is consumed. The circumstance, that an addition of iodine, hence a further strong augmentation of the absorption, does not influence the velocity, does not harmonise either with the assumption of the iodine-ion being the actual light-sensitive substance, but seems rather to point in the direction of the active co-operation of the iodine in the process.

But the most remarkable thing in PLOTNIKOW'S works is the experiments with the reaction-mixture as a light-filter. Even if the lack of effect of the iodine in the reaction-mixture itself can be understood, it must — when formed in the light-filter — absolutely absorb blue, violet and ultra-violet, and as the velocity is proportional to the intensity, the process in the reaction-vessel behind must necessarily be gradually retarded, notwithstanding its nature and which substance is to be regarded as the actual sensitive substance.

Methods and Results.

These various mysterious or improbable statements made me decide to take up the investigation of this process. To start with, the colour-sensitiveness was examined. The mea-

surements were already carried out in 1913 by Mr. Heegård-Poulsen. The arrangement is more closely described elsewhere[1]. It consisted materially in a quartz-mercury-lamp, a couple of screening shades, to procure a fairly parallel light, a filter-couvette, furnished with quartz-walls, and a reaction-couvette cemented on to it, without front-wall and with back-wall of quartz. The rate of reaction was measured by means of the energy-measurer, the fluormeter (which I have previously described[2]), the amounts of light passing through the filter and the reaction-mixture being measured during the progress of the process. When the extinction of the tri-iodine-ion for the various wave-lengths are known, its concentration and therewith the decrease in the concentration of the oxygen can be calculated at different times. As the process during the given circumstances occured very rapidly, energy-measurements had to be taken at such short intervals, that the time only allowed 1 adjustment each time, and the measurements cannot, therefore, lay claim to any great accuracy. But the question being here only of quite a rough determination of the colour sensitiveness, the method might apply.

The results showed that the transformed amount of substance per absorbed energy-unit was largest in violet (405 μμ) and somewhat smaller in blue (436 μμ) and ultra-violet (336 and 313 μμ).

The question was now allowed to rest until the spring of 1918. The preliminary experiments were executed by Mr. V. A. Larsen, the final measurements by me.

The necessary extinction-measurements were partly, for the lines 436 and 405 μμ, carried out by means of König-

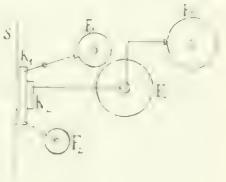
[1] Zeitschr. f. wiss. Photogr. **14**, 197 (1914).
[2] Zeitschr. f. Elektroch. **19**, 390 (1913)

Martens spectrophotometer, and partly, for the lines 366 and 313 μμ, with the fluormeter. The absorption of the iodine being very powerful and therefore only measurable with highly dilute solutions, the re-calculation factor becomes very great and the accuracy of the final figures consequently slight. The found extinction coefficients have the following values:

Table 1.

Konc. af *KJ*	Konc. af J_2	436	405	366	313 μμ
0,2—*n*	$5{,}75 \cdot 10^{-5}$	600	2100	6900	4150 -
1,0—*n*	„	900	2800	8050	4300 -
0,2—*n*	$11{,}50 \cdot 10^{-5}$	1000	2600	8600	3900 -
1,0—*n*	„	900	2500	9350	— -
		900	2500	8200	4100 μμ

In the insolation experiments I reasoned as follows: If Plotnikow's discovery, that the velocity is proportional to the oxygen-concentration, be correct, we must, by keeping the latter constant, be able to obtain constant velocity and thereby be able to investigate in the easiest and most accurate way the significance of the absorption of the tri-iodine-ion.



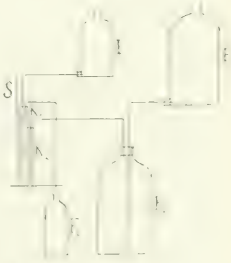


Fig. 1.

The applied arrangement was as schematisised in fig. 1.

L is the quartz-mercury-lamp, the electrode potential and current-strength of which was controlled all the time during the experiments and, where necessary, kept constant by variation of the inserted

resistance. S is a shade of card-board in which is an aperture of the same size as the reaction-couvette K_2. This as well as the large filter-couvette K_1 is made of glass having plan-parallel side-walls. Only in the experiments with the wave-length 313 μμ a couvette furnished with quartz-walls was used. The filter-liquid runs slowly from the flask F_1 down into the filter-couvette, and from there through a siphon over into the flask F_2. By this slow circulation of the filter-liquid we gain the advantage that the temperature in the reaction-couvette keeps very nearly constant. When the hydrochloric acid is furthermore heated so much as to just adopt the temperature of the room by the subsequent solution of potassium iodine, this temperature keeps sufficiently constant ($\pm 1°$) during the experiment. It must here be remembered, that the temperature-coefficient of the photolysis is so small (1,4, according to Plotnikow), that this variation is without any significance at all, and that the dark-process by ordinary temperature is so exceedingly slow, that its high temperature-coefficient plays no rôle whatever. The oxygen-concentration was kept constant by having a current of air bubbling through the reaction-mixture. The air was pressed out of the flask F_4 by means of water which ran down from the flask F_3. It was thus always saturated with water-vapour and was therefore unable to bring about any evaporation of the reaction-liquid, even if the current of air was very strong. As filters were employed those previously[1] described by me.

The final experiments took place, then, in the following manner. The lamp having reached constant tension and the filter-liquid having been put into slow circulation, the hydrochloric acid was heated (2—7° according to the

[1] Zeitschr. f. Electrochemie 19, 390 (1913).

concentration of the potassium iodide). In the darkened room, where all the measurements were carried out, the weighed amount of potassium iodide was then dissolved in hydrochloric acid by shaking, the temperature of the mixture measured and the couvette filled and put in its place. While this was going on the aperture in the shade S was covered with a shade of card-board. When now the air had been made to bubble through the reaction-mixture with the desired velocity, the aperture in the shade was uncovered and the time together with the voltage and current-strength of the lamp were noted. After the lapse of a suitable time a certain amount (as a rule 2 cm^3) of the reaction-mixture was taken out with a pipette, starch solution added and it was titrated with $^1/_{100}$-n thiosulphate of sodium. Immediately thereupon the voltage and the current-strength of the lamp were again read and, if necessary, regulated. After the majority of the liquid had been consumed in this manner, the couvette K_2 was removed, the fluormeter put up close behind the couvette K_1 and the incident energy measured, generally at various potentials of the normal lamp.

The provisional experiments now showed, in accordance with those of Mr. Heegård Poulsen, that a very good parallelism was really present between the absorbed amounts of energy of various wave-lengths and the corresponding amounts of substance transformed. But the provisional experiments having later turned out to have been made on false assumptions, I only shall mention them very briefly in order to obtain an opportunity of showing where the error came in. The velocity-curve in most of these experiments was as in fig. 2.

The velocity increased rapidly in the beginning, slower

later on, and ended by becoming fairly constant. In some of the experiments sufficient temperature constancy had not been provided for, the hydrochloric acid — having room-temperature — being much cooled by solving the potassium-iodide, and the temperature of the mixture therefore increased, first rapidly, later more slowly towards room-temperature. This source of error having been corrected, the velocity curve was directly from the beginning fairly rectilinear, as was to be expected. An addition of iodine did not alter the velocity, as was also found by Plotnikow.

Fig. 2.

This fact, in connection with the parallelism between the absorption of the tri-iodine-ion and the colour sensitiveness of the reaction-mixture which had been shown both by Heegård-Paulsen and especially by V. A. Larsen's experiments, made it probable to a high degree, that the tri-iodine-ion must be the actual light-sensitive substance, and that the course of the reaction-velocity might be computable when the amount of energy was known which in each single moment is absorbed by the tri-iodine-ion. If now we imagine a transverse section of there action-mixture at various times after the beginning of the exposure, it will be like fig. 3, where the inserted curves indicate the light-intensity in various depths of the liquid at the times t_1, t_2, etc. The simultaneously absorbed amounts of energy increase steadily, until at the

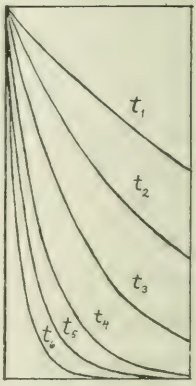


Fig. 3.

moment t_4, the absorption is complete, hence equal to 1. In this first phase of the exposure the reaction velocity, presupposed to be proportional to the absorbed amount of light, must constantly increase until the absorption becomes complete. From this moment the velocity must become constant.

When the exposure is continued the absorbed amount of light remains the same, but the exposed layer of the liquid is all the time decreasing in thickness. The process takes place in this exposed layer, and thereafter the reaction products are to be mixed with all the dark liquid-volume in the back part of the couvette. If the process is to advance with its full velocity, if, hence, the absorbed light is to be utilised as well as possible, this mixing-process — which may be accelerated by stirring — must take place so quickly, that even the velocity of the light-reaction in comparison herewith is negligible. Augmented stirring must, consequently, augment the rate until a certain limit.

The correctness of the theory can accordingly by tested in three ways. Partly by examining the occurrence of the reaction velocity in the beginning of the exposure, partly by examining the influence of the stirring, and finally by examining the relation between the maximum, constant, velocity and the simultaneously absorbed amount of energy of various wave-lengths.

The Initial Velocity.

The object being to get as many measurements as possible before the absorption becomes complete, these experiments were carried through for blue light of the wave-length 436 μμ, where the extinction-coefficient is comparatively slight (900), and with a couvette with only 5 mm

thickness of layer. The employed solution consisted of 3,32 gms. ($^1/_{50}$ mol) potassium-iodide which was dissolved in 20 cm^3. 1-*n* hydrochloric acid.

In the following tables:

t means the time, expressed in minutes.

x the liberated amount of iodine, in equivalents per liter.

Δx the increase in the amount of iodine.

$1—10^{-\frac{\alpha}{2}x}$ the relative absorption.

$\Delta x : \left(1—10^{-\frac{\alpha}{2}x}\right)$ the transformed amount of substance per unit of absorption. Δx corresponds to the period which comes after each single determination of absorption.

Table 2.

Experiment 1.

Distance 30 cm. 340 Watt. $T = 26°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$1—10^{-\frac{\alpha}{2}x}$	$\Delta x \cdot 10^3 : \left(1—10^{-\frac{\alpha}{2}x}\right)$
2	0	—	—	—
4	0	—	—	—
6	0	—	—	—
9	0,28	—	0,30	—
12	0,35	0,07	0,36	0,23
15	0,53	0,18	0,49	0,50
18	0,71	0,18	0,59	0,37
21	0,95	0,24	0,70	0,41
				0,38

$$\frac{\Delta x \cdot 10^3}{1—10^{-\frac{\alpha}{2}x}} : \Delta t = 0{,}13$$

Experiment 2.

Distance 30 cm. 350 Watt. $T = 26°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$1—10^{-\frac{\alpha}{2}x}$	$\Delta x \cdot 10^3 : \left(1—10^{-\frac{\alpha}{2}x}\right)$
4	0	—	—	—
8	0	—	—	—
12	0	—	—	—
16	0,19	—	0,21	—
20	0,38	0,19	0,38	0,90

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$1-10^{-\frac{\alpha}{2}x}$	$\Delta x \cdot 10^3 : \left(1-10^{-\frac{\alpha}{2}x}\right)$
24	0,60	0,22	0,53	0,58
28	0,95	0,35	0,70	0,66
32	1,25	0,30	0,79	0,43
				0,64

$$\frac{\Delta x \cdot 10^3}{1-10^{-\frac{\alpha}{2}x}} : \Delta t = 0{,}16$$

Experiment 3.

Distance 30 cm. 390 Watt. $T = 25°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$1-10^{-\frac{\alpha}{2}x}$	$\Delta x \cdot 10^3 : \left(1-10^{-\frac{\alpha}{2}x}\right)$
10	0,65	—	0,56	—
14	0,90	0,25	0,68	0,45
18	1,28	0,38	0,80	0,56
22	1,70	0,42	0,89	0,52
26	1,95	0,25	0,92	0,28
30	2,21	0,26	0,94	0,28
34	2,85	0,64	0,98	0,68
38	3,23	0,38	0,99	0,39
42	3,63	0,40	0,99	0,40
				0,45

$$\frac{\Delta x \cdot 10^3}{1-10^{-\frac{\alpha}{2}x}} : \Delta t = 0{,}11$$

Experiment 4.

Distance 30 cm. 360 Watt. $T = 26°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$1-10^{-\frac{\alpha}{2}x}$	$\Delta x \cdot 10^3 : \left(1-10^{-\frac{\alpha}{2}x}\right)$
22	0	—	—	—
30	0,29	—	0,30	—
34	0,65	0,36	0,56	1,2
38	0,81	0,16	0,64	0,29
42	0,98	0,17	0,71	0,27
46	1,39	0,41	0,83	0,58
50	1,65	0,26	0,88	0,31
54	2,00	0,35	0,92	0,40
58	2,38	0,38	0,95	0,41
				0,49

$$\frac{\Delta x \cdot 10^3}{1-10^{-\frac{\alpha}{2}x}} : \Delta t = 0.12$$

Experiment 5.
Distance 60 cm. 310 Watt. $T = 26°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$1 - 10^{-\frac{\alpha}{2}x}$	$\Delta x \cdot 10^3 : \left(1 - 10^{-\frac{\alpha}{2}x}\right)$
0	0,11	—	0,13	—
10	0,39	0,28	0,39	(2,20)
20	0,66	0,27	0,56	0,69
30	1,04	0,38	0,73	0,68
40	1,46	0,42	0,84	0,58
50	1,75	0,29	0,89	0,35
60	2,23	0,48	0,94	0,54
70	2,70	0,47	0,97	0,50
80	2,98	0,28	0,98	0,29
				0,52

$$\frac{\Delta x \cdot 10^3}{1 - 10^{-\frac{\alpha}{2}x}} : \Delta t = 0{,}052$$

The four first experiments, where the lamp was placed in 30 centimeters' distance from the reaction-couvette, give the following amounts of substance transformed per unit of absorption and time:

Table 3.

Experiment	$\frac{\Delta x \cdot 10^3}{1 - 10^{-\frac{\alpha}{2}x}} : \Delta t$
1	0,13
2	0,16
3	0,11
4	0,12
	0,13

Hence the velocity formula becomes:

$$\frac{\Delta x \cdot 10^3}{1 - 10^{-\frac{\alpha}{2}x}} = 0{,}13\,\Delta t.$$

By means of this formula the drawn curve in fig. 4 is calculated, and the observed points are entered here upon, each row of points being displaced in horizontal direction,

until the mean deviations from the curve are as small as possible. The zero-point being uncertain (it depends upon the degree of purity of the sample of potassium iodide) this way of proceeding is fully justifiable.

The figure and the tables show the following facts. For the experiments 1 and 2 a very pure sample of potassium iodide was employed, which even in acid liquid and with admittance of air, kept constant for a long time without giving colouring with starch. Even by exposure a measur-

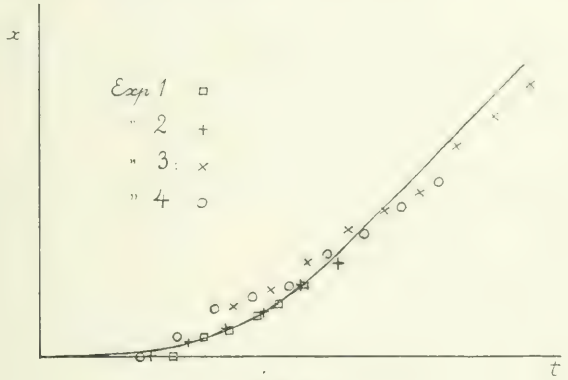


Fig. 4.

able iodine-separation does not occur until after 9—16 minutes respectively, but it then continues with increasing velocity, exactly as demanded by the theory. In these cases the process consequently shows both zero-period and induction-period, — contrary to PLOTNIKOW's assertion. All the induction-periods, the cause of which we have hitherto been able to trace, originate, however, in the fact that some inhibiting impurity has had to be removed before the process can be properly set going. This is, f. inst. the case with the mixture of chlorine and hydrogen, where the impurity is not indeed known with certainty,

but is supposed to be ammonia or an ammonia-derivative[1], with photochlorination of benzene and similar substances, where the impurity is oxygen[2], and with Eder's liquid, where it is likewise oxygen which is removed in the course of the induction-period[3].

But in the case here investigated the induction-period, on the contrary, arises from the fact, that the actual light sensitive substance, the tri-iodine-ion must first be procured before the process can proceed by means of the energy absorbed by this light sensitive substance. The first small amount of tri-iodine-ion cannot, accordingly, be procured by the light itself, as the effect of the light only begins when absorbing tri-iodine-ion is present. It must be secured by means of the slow oxidation, which also takes place in the dark, and we thereby reach the peculiar conclusion, that the condition of processes of this kind being light sensitive is their possession of a measurable darkness-velocity. Thereby they form a separate group, as will be more closely developed in another section (v. p. 24).

If now the employed sample of potassium-iodide contains a little iodate or another oxidating impurity which immediately by solving in hydrochloric acid liberates a little iodine, the zero-period will be quite omitted, and the induction-period be materially shortened, the light immediately finding tri-iodine-ion to work with. This was the case in the experiments 3 and 4, where the employed potassium iodide immediately after the solution in hydrochloride acid gave colouring with starch.

[1] Burgess and Chapman, Journ. Chem. Soc. **89**, 1402 (1906).
[2] Luther and Goldberg, Zeitschr. für phys. Chemi **56**, 43 (1906).
[3] Winther, Zeitschr. f. wiss. Photogr. **8**, 197 (1910).

The Effect of Stirring.

Already the form of the velocity curve in the provisional experiments showed that the slow bubbling-through of air which was employed, produced a too feeble stirring to make the mixing-process occur very rapidly in comparison with the light-process. This is corroborated by the three following experiments which were made under the same circumstances as the experiments 1—4, with the difference, only, that the air-current in the last named experiments was far stronger than in the experiments 6—8.

Table 4.

$\lambda = 436$ μμ. Thickness of layer $^1/_2$ cm. Distance 30 cm.

Experiment 6.

360 Watt. $T = 26°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$1 - 10^{-\frac{\alpha}{2}x}$	$\Delta x \cdot 10^3 : \left(1 - 10^{-\frac{\alpha}{2}x}\right)$
6	0	—	—	—
12	0	—	—	—
18	0	—	—	—
24	0	—	—	—
30	0,28	—	0,30	—
40	0,78	0,50	0,63	1,67
50	1,34	0,56	0,82	0,89
60	1,93	0,59	0,91	0,72
70	2,56	0,63	0,96	0,69

Experiment 7.

340 Watt. $T = 26°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$1 - 10^{-\frac{\alpha}{2}x}$	$\Delta x \cdot 10^3 : \left(1 - 10^{-\frac{\alpha}{2}x}\right)$
16	0	—	—	—
24	0	—	—	—
30	0,25	—	0,27	—
36	0,55	0,30	0,50	1,11
42	0,90	0,35	0,68	0,70
48	1,20	0,30	0,78	0,44
54	1,50	0,30	0,85	0,38
60	1,73	0,23	0,89	0,27
66	2,15	0,42	0,94	0,47

Experiment 8.
340 Watt. $T = 26°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$1-10^{-\frac{\alpha}{2}x}$	$\Delta x \cdot 10^3 : \left(1-10^{-\frac{\alpha}{2}x}\right)$
18	0,96	—	0,70	—
22	1,29	0,33	0,81	0,47
26	1,65	0,36	0,88	0,44
30	1,88	0,23	0,91	0,26
34	2,06	0,18	0,93	0,20
38	2,26	0,20	0,95	0,22
42	2,50	0,24	0,96	0,25

The tables show, that the quotient $\Delta x : \left(1-10^{-\frac{\alpha}{2}x}\right)$ i. e. the transformed amount of substance per absorbed unit of energy, decreases greatly when the absorption increases, and that it is almost everywhere smaller than it ought to be according to the experiments 1—4. In experiment 6 it ought to be $10 \cdot 0{,}12 = 1{,}20$, in experiment 7, 0,72 and in experiment 8, 0,48. The slow stirring has accordingly prevented the absorbed light from being utilised to its full degree.

The same result is attained by a couple of experiments, in which the stirring-velocity was altered during the experiment. The further details having no interest in this place I shall only give the transformed amounts of substances. The wave-length was in both experiments 366 μμ.

Table 5.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	t	$x \cdot 10^3$	$\Delta x \cdot 10^3$
	strong stirring			strong stirring	
6	0,50	—	6	0,58	—
12	1,13	0,63	12	1,12	0,54
18	1,63	0,50		feeble stirring	
24	2,03	0,40	18	1,30	0,18
	feeble stirring			strong stirring	
30	2,35	0,32	24	1,75	0,45
	strong stirring		30	2,10	0,35
36	2,88	0,53			
	feeble stirring				
42	3,19	0,31			

A comparison between the provisional and the final experiments finally showed, that the constant velocities which were attained in the former, are everywhere much smaller than in the same circumstances in the latter, where the stirring was stronger.

The Colour Sensitiveness.

For the wave-lengths 405, 366 and 313 μμ, the extinction is so strong that the absorption even with very small amounts of iodine is practically complete. With regard to these wave-lengths there could therefore be no question of examining the induction-period as in the case of 436 μμ, but we had to keep to the constant velocity which occurs in complete absorption. In order to obtain constant velocity also with the blue line (436 μμ) without exposing too long, so much iodine-solution was in this case previously added as to make the absorption very nearly complete at the beginning of the exposure. In the experiments 9—13, 0,5 cm thickness of layer was employed and a lamp-distance of 30 cm. The solution was, as before, normal, both with regard to hydrochloric acid and potassium iodide.

Table 6.

Experiment 9.

$\lambda = 436$ μμ. 365 Watt. $T = 26°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$\Delta x \cdot 10^3 : \Delta t$
0	2,75	—	—
4	3,10	0,35	0,09
8	3,50	0,40	0,10
12	3,78	0,28	0,07
16	4,19	0,41	0,10
24	4,93	0,74	0,09
32	5,55	0,62	0,08
40	6,38	0,83	0,10
			0,090

Experiment 10.

$\lambda = 405$ μμ. 360 Watt. $T = 26°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$\Delta x \cdot 10^3 : \Delta t$
8	0,75	—	—
14	1,23	0,48	0,08
20	1,59	0,36	0,06
26	2,00	0,41	0,07
32	2,44	0,44	0,07
38	2,85	0,41	0,07
44	3,34	0,49	0,08
50	3,75	0,41	0,07
			0,071

Experiment 11.

$\lambda = 405$ µµ. 345 Watt. $T = 24°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$\Delta x \cdot 10^3 : \Delta t$
6	0,56	—	—
12	0,94	0,38	0,06
18	1,38	0,44	0,07
24	1,69	0,31	0,05
30	2,00	0,31	0,05
36	2,41	0,41	0,07
42	2,78	0,37	0,06
			0,060

Experiment 12.

$\lambda = 366$ µµ. 405 Watt. $T = 25°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$\Delta x \cdot 10^3 : \Delta t$
6	0,50	—	—
12	1,13	0,63	0,11
18	1,63	0,50	0,08
24	2,03	0,40	0,07
30	2,56	0,53	0,09
36	2,90	0,34	0,06
			0,082

Experiment 13.

$\lambda = 366$ µµ. 335 Watt. $T = 26°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$\Delta x \cdot 10^3 : t$
6	0,58	—	—
12	1,13	0,55	0,09
18	1,46	0,33	0,06
24	1,81	0,35	0,06
30	2,08	0,27	0,05
36	2,42	0,34	0,06
42	2,83	0,41	0,07
48	3,11	0,28	0,05
			0,063

Experiment 14.

$\lambda = 313$ µµ. 360 Watt. $T = 25°$, thickness of layer 1 cm.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$\Delta x \cdot 10^3 : \Delta t$
10	0,81	—	—
20	1,56	0,75	0,08
30	2,38	0,82	0,08
			0,080

Experiment 15.

$\lambda = 313$ µµ. 360 Watt. $T = 27°$, thickness of layer 1 cm.

t	$x \cdot 10^3$	$x \cdot 10^3$	$x \cdot 10^3 : t$
10	0.75	—	—
20	1,63	0,88	0,09
30	2,50	0,87	0,09
			0,090

These experiments show that the velocity — by complete absorption and constant oxygen-concentration — actually is as nearly constant as can be expected with the relatively big experimental errors which are inevitable in photochemical measurements of this kind, where it is important to work in as thin layers as possible, and where the amounts of substances which are to be measured must consequently be very small.

Table 7 contains the found velocities together with the simultaneously measured absorbed amounts of energy (in erg per mm^2 and second) and the quotients of these figures.

Table 7.

λ	$\Delta x \cdot 10^3 : \Delta t$	$E_{abs.}$	$\frac{\Delta x \cdot 10^5}{\Delta t} : E_{abs}$
436	0,090	5,9	1,5
405	0,071	2,9	2,5
—	0,060	2,3	2,6
366	0,082	5,5	1,5
—	0,063	3,5	1,8
313	0,080	3,1	2,6
—	0,090	3,5	2,6
			2,2

When considering the inaccuracy which an energy-measurement with the fluormeter is always beset with, the proportionality between the absorbed amounts of energy and the transformed amounts of substance can hardly be expected to be shown in a better manner.

According to PLOTNIKOW's experiments, which were, however, carried out under quite different circumstances — no stirring, decreasing oxygen-concentration — the velocity should be proportional to the oxygen-concentration, for which reason a single series of experiments was executed with the passage of pure oxygen. It hereby appeared that the darkness-velocity increased more than the oxygen-concentration. While the darkness-velocity in the former experiments on an average is $0{,}11 \cdot 10^{-3}$ equivalents per liter in the course of 15 minutes, by saturating with oxygen it was found to be $0{,}66 \cdot 10^{-3}$. The light-velocity, on the contrary, increased in far slighter degree. In order to be able to eliminate the great darkness-velocity the experiment was carried through with alternating exposing and darkening.

Table 8.

$\lambda = 366$. 350 Watt. $T = 25°$.

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$\Delta x \cdot 10^3 : \Delta t$
0	0	—	—
darkness			
6	0,25	0,25	—

t	$x \cdot 10^3$	$\Delta x \cdot 10^3$	$\Delta x \cdot 10^3 : \Delta t$
light			
12	1,91	1,66	0,23
darkness			
18	2,13	0.22	—
light			
24	3,46	1,33	0,17
darkness			
30	3,79	0,33	—
light			
36	4,90	1,11	0,15
			0,18

The mean of the darkness-velocities is $0{,}27 \cdot 10^{-3}$ per 6 minutes. If this velocity is subtracted from the velocity of the total amounts transformed, we obtain the light-velocities per minute, given in the last column. By comparison with the experiments 12 and 13 it is seen, that the velocity has only increased abt. $2^1/_2$ times, simultaneously with the oxygen-concentration having been augmented 5 times. The absorbed energy here being 3,7 erg, we have:

$$\frac{\Delta x \cdot 10^5}{\Delta t} : E_{\text{abs}} = 4{,}9$$

2,2 times as great, then, as the average for the experiments with air-passage.

Before leaving this section we have only to note, that Plotnikow's assertion, that the reaction-mixture employed as a light-filter, does not decrease the velocity, must either be due to a misunderstanding, an incorrect statement in the said paper, or to quite incredible experimental errors.

For the investigation of this point a couvette of 1 cm thickness was divided into two partitions by means of a thin glass-wall, each partition of abt. 0,45 cm thickness, and both filled with the usual reaction-mixture and stirred by an aircurrent, the stirring — which was insufficient for a complete utilising of the light — being as nearly uniform

in both halves as possible. The light had consequently to pass the one partition with the reaction-mixture before it reached into the other partition. At the same time the velocity of the dark-process was determined.

Table 9.

$\lambda = 405$ µµ. 350 Watt. $T = 26°$.

t	$x \cdot 10^3$ 1st partition	2nd partition	darkness	2nd part. ÷ darkness
15	—	0,73	0,12	0,61
17	1,04	—	—	—
31	—	1,00	0,24	0,76
33	1,75	—	—	—
45	—	1,13	0,33	0,80
47	2,44	—	—	
60	—	1,25	0,45	0,80
62	3,23	—	—	—
75	—	1,46	0,57	0,89
77	3,88	—	—	—
90	—	1,58	0,69	0,89
92	4,55	—	—	—
105	—	1,75	0,78	0,97
107	5,19	—	—	—

While the process in the first partition takes its even course, it soon stops in the second partition, of course on account of the absorption of the iodine in the reaction-mixture standing before it. To make certain that this stop in the 2nd partition was not due to peculiarities in the walls of this partition or in the medium wall, it was filled, in a fresh experiment, with the same reaction-mixture, but in such a way that the 1st partition was filled with 1 normal hydrochloric-acid.

Table 10.

$\lambda = 405$ µµ. 420 Watt. $T = 26°$.

t	$x \cdot 10^3$
15	1,10
30	2,36
47	3,75
60	4,69

By comparison with the last experiment it is seen that the 2nd partition acts quite normally, if only the active light is not kept away.

II. The Optical Auto-Sensitisation.

The previous investigations show that, by the photochemical oxidation of the iodine-hydrogen, it is the tri-iodine-ions, themselves formed by the process, which determine the course of the process, the amount of energy, absorbed by these tri-iodin-ions, being decisive for the reaction-velocity. The absorption of the iodine-ion lies further out in ultra-violet. The figures 5 and 6 show the course of the absorption by the formation of the tri-iodine-ion. They are the result of spectrographical investigation partly of solutions of a certain amount of iodine in varying concentrations of potassium-iodide (fig. 5), partly of varying amounts of iodine in 1-normal potassium-iodide-solution (fig. 6). In fig. 5 are furthermore inserted the absorption-limits for varying concentrations of potassium-iodide. It is seen from this that the absorption of the potassium-iodide only becomes strong at abt. 280 $\mu\mu$, while the absorption of the tri-iodine-ion reaches into the visible spectrum, forming two bands having maxima at abt. 350 and abt. 290, the latter being, however, deformed on account of the absorption of the potassium-iodide in this region. Fig. 5 shows how the absorption of the solution decreases greatly with falling potassium-iodide-concentration, hence, with decreasing tri-iodide-concentration, the tri-iodide, thus, being undoubtedly the bearer of the strong absorption.

The energy absorbed by the tri-iodine-ion must in one way or other be transmittable to the iodine-ion, and thereby

give rise to the oxidation of the latter. As will be shown in another paper, the question here, probably, is of a transformation of the absorbed radiation to a radiation of such wave-length as can be absorbed by the iodine-ion.

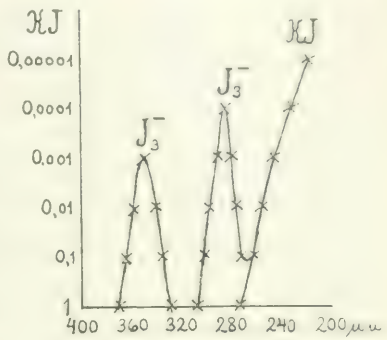


Fig. 5.

In this place we have only to point out what is essential in the process, namely 1) that the optical sensitiser is first formed in the process itself, which consequently 2) must possess a measurable darkness-velocity, in order that this sensitiser can be formed and act at all, and 3) that the optical sensitiser is formed by simple oxidation of the substance which is to be oxidated by the process, hence by the process itself which it is to accelerate.

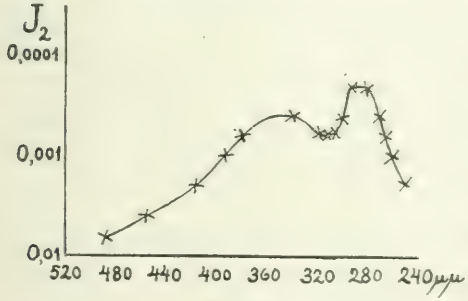


Fig. 6.

It is now beyond all doubt that these peculiarities are refound in a series of other processes and are so characteristic that we are fully justified in comprising these processes in a group which may suitably be called the optical auto-sensitisation[1].

[1] I. Plotnikow in his photochemical system (Zeitschr. f. physikal. Chemie 77, 472) has a group which is called »photochemische Autokatalyse« which should approximately approach the one given here, but wanting in several of the essential features of the latter, and without statement of any example.

In the literature we find the following cases:

The oxidation of the leuco-compounds to colour-substances. This process possesses a measurable darkness-velocity. It is, as shown by Gros[1], light-sensitive and is — as shown by the same author — accelerated by the colour substances themselves which are formed by the oxidation, the rays which are absorbed by these colour-substances being thus those which accelerate the oxidation also of the, to begin with, (apparently) colourless leuco-compounds. Starting from the indicated theory of the transformation of the absorbed radiation, it is of interest to note, that the leuco-compounds and the corresponding colour-substances according to P. Krüss[2] possess a common absorption-domain in ultra-violet, quite like iodine-ion and tri-iodine-ion.

The oxidation of the pyrogallic acid in alcaline liquid. This process, according to Trautz and Thomas[3], is sensitive to red light, which is quite mysterious as long as the pyrogallic acid itself is considered as the actual sensitive substance. The oxidation product is brown. Its extinction-curve Mr. Orla Saebye has determined with König-Martens spectrophotometer. The solution contained 2 cm^3 5 per cent pyrogallic acid, 2 cm^3 10 per cent sodium hydroxide, 300 cm^3 water, and was oxidated by shaking with air. The concentration of the pyrogallic acid was 0,00262-molar. The molar decadic extinction-coefficients are given in table 11 and fig. 7.

Table 11.

Solution:

I			II		
wave-length	fresh	1 day old	wave-length	fresh	1 day old
694 μμ	—	65	692 μμ	90,5	47,5
656 -	—	8,7	667 -	103	0

[1] Zeitschr. f. physikal. Chemie 37, 157 (1901).

[2] Jahrbuch d. Photographie 1906, 83.

[3] Thomas, Inaug, Dissert. Freiburg i. Br. 1908, 39.

Solution:

I			II	
wave-length	fresh	1 day old	wave-length	fresh	1 day old
579 μμ	220	23,1	646 μμ	144	3,2
546 -	257	105	626 -	192	22
436 -	404	591	607 -	242	77
			579 -	340	117
			546 -	445	232
			436 -	1185	—

It is seen from this, that by simple oxidation of pyrogallic acid at any rate there may be formed a product which has selective absorption in red. Whether it is just that substance which is formed by the insolation of pyrogallic acid must be determined by direct experiments, but the sensitiveness has, however, been found to be so small that even a slight amount of this substance is sufficient to explain it.

Fig. 7.

Also this process has a measurable darkness-velocity.

The oxidation of sodium sulphide in aqueous solution. The process possesses a measurable darkness-velocity. According to Trautz[1] it is not light-sensitive, as long as the solution is pure, but becomes so when by oxidation free sulphur (or polysulphides) are formed. The sensitiveness lying also here in red, it must be specially investigated whether the light-product also here has a selective absorption in this spectral-section.

Provisionally we thus know four photolyses which

[1] Jahrb. d. Photographie 1909, 39.

undoubtedly fall within this group, and several other light-processes will no doubt on closer examination be referable to it.

Results of this work.

When solutions of hydriodic acid are insolated in thin layer, strongly stirred and kept at a constant concentration of oxygen, the rate of oxidation is at first zero, then increases during an induction period and finally becomes constant.

The sensitiveness of the oxidation of hydriodic acid is — both for the same wavelength and different concentrations and for different wavelengths — proportional to the absorption of the triiodine-ion.

It is possible to form a certain group of optical sensitisations — the optical autosensitisation — the characteristics of which are, that the optical sensitiser is formed by the very process which it is to accelerate, and that this process therefore must possess a measurable velocity in darkness.

Royal Technical College of Copenhagen,
May 1919.

Færdig fra Trykkeriet den 10. Februar 1920.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. II, 3.

THE PHOTOCHEMICAL EFFICIENCY OF THE ABSORBED RADIATION

BY

CHR. WINTHER

KØBENHAVN

HOVEDKOMMISSIONÆR: ANDR. FRED. HOST & SØN, KGL. HOF-BOGHANDEL

BIANCO LUNOS BOGTRYKKERI

1920

I. The Efficiency of the Absorbed Radiation.

Generally speaking the absorbed radiation will — perhaps through several intermediate links — finally be transformed to thermal energy. In those cases in which a chemical process takes place in the absorbing system in consequence of the absorption, part of the absorbed radiation must, — as has long been known[1] — at any rate provisionally, be transformed to chemical energy. One of the most important problems in modern photochemistry has accordingly been, to find relations between the absorbed amount of energy and the chemical change produced hereby, in other words, to find a regularity in the numerical expressions for the light-sensitiveness which apparently reaches from very great to very small values. The task may be still more accurately defined.

It has long been customary, according to Helmholz'[2] example, to divide the photochemical processes into reversible or work-accumulating and irreversible or work-yielding, after the analogy of the division of the darkness-processes. This division is, as far I can see, erroneous, as it does not at all touch upon the actual photochemical part of the total light-process, which must in all cases be considered as complex. This actual photo-

[1] Ostwald, Lehrb. d. allgem. Chemie II. 1, 1087.

[2] Erhaltung d. Kraft 25 (1847).

chemical division-process, the primary photochemical process, consists — like ordinary light-absorption — in an absorption of radiation-energy, i. e. in an augmentation of the chemical potential of the absorbing substance. After this follows one or more, necessarily voluntary, secondary darkness processes which, hence, occur with loss of free energy. These two processes being quite independent of each other, it is evidently quite indifferent to the primary process whether this secondary loss of free energy is greater or smaller than the primary augmentation of the same. In the first case we have an irreversible, in the second a reversible total light-process, and the argument shows that, with regard to the photochemical part of the process, there is no difference between the two groups. For the total process the difference does not arise until the moment when the reverse reaction in the reversible process under the given circumstances has a measurable velocity.

Photochemical statics and dynamics render us several examples of the correctness of this reasoning.

As an example Warburg[1], Regener[2] and Coehn[3] have examined the decomposition of ammonia into nitrogen and hydrogen at ultra-violet insolation. By combining the observations of these investigators the velocity of decomposition proves constant until the decomposition is complete. At ordinary temperature and pressure in darkness the equilibrium lies at abt. 1.5 per cent decomposition. If pure ammonia, there, be exposed, the radiation acts in the same direction as the chemical forces in the first

[1] Sitzungsber. d. Berliner-Akad. 1911, 746.
[2] Sitzungsber. d. Berliner-Akad. 1904, 1228, Drud. Ann. 20, 1033 (1906).
[3] Jahrb. Rad. Elektr. 7. 615 (1910), Zeitschr. f. Elektroch. 20, 275 (1914).

$1^1/_2$ per cent decomposition, the process is, consequently, irreversible. As soon as we reach the $1^1/_2$ per cent decomposition, the process becomes an equilibrium displacement, hence reversible, without the velocity in any way being affected thereby. The fact that the decomposition can here be complete is partly due to the circumstance that the reverse process, the formation of ammonia, has no measurable velocity at ordinary temperature, partly to its lack of sensitiveness to the radiation employed.

Quite similarly COEHN and BECKER[1] have been able to displace the equilibrium $SO_2 + O \rightleftarrows SO_3$ at ultra-violet insolation without it having been noticeable in the velocity when the darkness-equilibrium was passed. When here only a SO_3 decomposition of 35 per cent was reached, in spite of the darkness-velocity of the process under the given circumstances being exceedingly slight, it is due to the fact, that both processes are here sensitive to the employed, mixed radiation, and that the obtained stationary state consequently depends upon the spectral composition of the radiation.

As a third example may be named the process $HgCl_2 + FeCl_2 \rightleftarrows HgCl + FeCl_3$, which I have previously[2] investigated. When the left hand system was exposed in the presence of oxygen, a general oxidation of the ferro-salt to basic ferri-salt simultaneously took place, which process is sensitive to the same spectral region as the first named, so that in both cases it is undoubtedly the ferro-salt, which is the actual sensitive substance. The distribution of ferro-salt between these two processes is now exclusively dependent on the ratio between the amounts of mercury-

[1] Zeitschr. f. physikal. Chemie. 70, 88 (1909).

[2] Overs. o. d. kgl. danske Vid. Selsk. Forh. 1911, Nr. 5.

salt and oxygen, and of the velocity by which exposed ferro-salt reacts with these substances, but without regard to the fact of the one process being reversible, the other irreversible.

Finally we shall mention that Luther and Weigert[1] by exposure of anthracene-solution formed measurable quantities of dianthracene, which was in inverse proportion to the velocity of the opposite transformation, and that Weigert[2] at 500° was unable to obtain any displacement at all of the equilibrium $COCl_2 \rightleftarrows CO + Cl_2$, despite the fact of both the reverse processes being sensitive. The cause is that both processes at the high temperature occur too rapidly for the light to compete with it.

These examples, to which several others could be added, show, conformably to the above reasoning, that the effect of the exposure is not at all dependent upon the magnitude and direction of the affinity in the total process, but that, in the so-called reversible light-processes, the question is solely of a competition between the light-effect and the velocity of the reverse process.

In spite of this, it will be practical provisionally to keep to the above division also here.

For the reversible processes, in which the total light-effect consists in an augmentation of free energy, it will be natural to compare this augmentation of energy with the absorbed energy of radiation which has had to be employed in order to produce it. In this manner several investigators, amongst whom especially Weigert[3], have computed the efficiency for various reversible processes and

[1] Zeitschr. f. physikal. Chemie 51, 297. 53, 385 (1905).
[2] Ann. d. Physik (4) 24, 55 (1907).
[3] Zeitschr f. wiss. Photogr. 11, 381 (1912).

thereby found figures which vary from some per mille up to 100 per cent (for the carbonic-acid-assimilation). In these calculations the total augmentation of energy has most frequently been employed, but — as shown by WEIGERT — it would have been more correct to reckon with the increase in free energy, the work effected by the radiation being only computable thereby. In consequence of the above reasoning, this way of calculation can only be of technical and not of special photochemical interest, as it includes the alteration of energy in the secondary process which — for the same primary process, hence the same effect of the absorbed radiation — may involve highly different displacements of energy.

The calculations made according to the same principle on irreversible processes possess still slighter direct significance. The question here not being of accumulation, either of total or free energy, we have had to restrict ourselves to putting the absorbed amount of energy in relation to the transformation of substance produced thereby, or to the loss in total or free energy corresponding to this transformation. Thus WARBURG[1] in a series of papers has determined the "specific photochemical effect", by which he understands the quantity of gr. molecules transformed per absorbed calory. It will be shown below how this quantity on certain conditions may be of unsuspected importance.

It is evident from the above explanation that the more insight we get into the character of the primary photochemical process, the more exactly shall we be able to determine the efficiency of the absorbed radiation. Among

[1] Sitzungsber. d. Berliner-Akad. 1911, 746. 1912, 216. 1913, 644. 1914, 872. 1915, 230. 1916, 314. 1918, 300.

the manifold hypotheses put forth to this end, only a single one has hitherto been applicable as a foundation for quantitative calculation. This hypothesis takes it for granted that the primary photochemical effect consists in a split off of electrons produced by the effect of resonance from the absorbed radiation, the frequency of which must fairly coincide with the "characteristic vibration" of the electron. The quantitative forming of this view is due to Einstein[1], who in 1905 advanced his "photochemical law of equivalence", according to which, in simple cases, each absorbed quantum can liberate 1 electron. If the said process be of so simple a nature as to enable us to judge of the magnitude of the transformation which will be the result of the split off of 1 electron, then we have here a possibility of calculating beforehand the total chemical effect of a certain absorption, when the wave-length, and with it the magnitude of the quantum, is known. This calculation Warburg has tried to carry through in the above mentioned series of papers, whereby he found good accordance with the experiments in some cases and not in others. But it is to be remembered, that everyone of these kind of calculations requires a special hypothesis of the mechanism of the said process.

Without knowing Einstein's work I tried in 1912[2] to compare the absorbed amounts of light with the amounts of energy which according to the quantum-theory are — for a series of light-processes — necessary for the split off of an amount of electrons, corresponding to the observed transformation of substance. By means of the information at hand in the literature I attained to a system of the

[1] Ann. d. Physik (4) 17, 148 (1905). 37, 832 (1912).
[2] Zeitschr. f. wiss. Phot. 11, 92 (1912).

photochemical processes founded on energetic basis. Others[1] have later made similar measurements and calculations, and in 1913 BODENSTEIN[2] on the same basis, but without knowing my work, has set forth a detailed theory of the nature of the photochemical effect. All these works have shown that, as a rule, there is no simple relation between the absorbed energy and the one required by the quantum-theory, the latter being very frequently far greater than the former.

The assumption, that the primary photochemical process consists in a split off of electrons, has naturally given rise to a series of attempts to demonstrate these liberated electrons, which result in a demonstration of light-electrical sensitiveness (Hallwachs-Effect) in the primary sensitive substance. All these experiments[3] gave a negative result. This, in connection with the fact that the absorbed amounts of energy in many cases are too small too explain the necessary split off of electrons, must — according to my opinion — cause us to disregard this explanation, at any rate as a general théory. By the atom-model made by N. BOHR[4], according to which the Planck-effect-quantum $h\nu$ no longer corresponds to the complete split off of electrons, but only to the moving of these from one stationary state to another, we should presumably get

[1] HENRI and WURMSER, Compt. rend. 156, 1012 (1913). 157, 126 (1913). BOLL, Compt. rend. 156, 1891 (1913). L'évolution photochimique des électrolytes, Paris 1914.

[2] Zeitschr. f. physikal. Chemie 85, 329 (1913).

[3] KUMMEL, Zeitschr. f. Elektroch. 17, 409 (1911). LUDLAM, Philos. Mag. (6), 23, 757 (1912). LENARD and RAMSAUER, Sitzungsber. Heidelb. Akad. 1910, Nr. 28, 31, 32, 1911, Nr. 16, 24. VOLMER, Ann. d. Phys. (4) 40, 775 (1913). LE BLANC and VOLMER, Zeitschr. f. Elektroch. 20, 494 (1914).

[4] Phil. Mag. (6), 25, 10 (1913).

over the difficulty which lies in the lacking light-electrical sensitiveness in the primary sensitive substances. But the other difficulty, that the absorbed amounts of energy are too small for the moving of the electrons to an exterior stationary condition, nevertheless remains.

On the other hand it looks as if — starting from quite another consideration — a quantitative accordance between the absorbed and the necessary amounts of energy can be demonstrated. As has first been pointed out by Arrhenius[1], a reaction with measurable velocity can only be imagined possible if a certain, slight fraction of the molecules at hand at each single moment are in such a state that they are able to react, hence, are active. If they were all in the same state, they would be obliged, either to react all at the same time, or not at all. The velocity of reaction under given circumstances must according to this conception be a measure for the fraction of all of the molecules in the active state. I have previously[2] applied this view for an explanation of the optical sensitisation. In accordance with this, the effect of the sensitiser should consist in transforming the absorbed radiation in such a way, as to render the energy gained thereby able to give some of the other molecules present that accretion of energy which would make them active, "help them over the top".

R. Marcelin[3] has already in 1914, however, been able to convert these purely qualitative considerations into quantitative calculations.

Marcelin compares the distribution of energy for a

[1] Bihang till kgl. Svenska Vet. Akad. Handl. 8, No. 13, 14 (1884).
[2] Forh. 16. skand. Naturforskerm. 1916, 279.
[3] Ann. de physique 3, 120 (1914).

large amount of molecules with a swarm of bees gathering around the queen. The swarm is densest towards the centre and decreases in density outward, the distribution being thus determined by the MAXWELL-BOLTZMANN-GIBBS-distribution-law. If now we imagine a surface of a sphere at a great distance from the centre, then in each single moment only quite a few of the bees will pass this surface from within and outward. If now these bees be supposed to be removed as soon as they pass the surface of the sphere from within, then, next moment, they will be replaced by others, which are also removed, and so on, and the whole phenomenon then gives the picture of a slowly occurring process. This sphere-surface represents a certain content of energy per gr. molecule, which MARCELIN calls the "critical energy", and the bees which reach the surface correspond to the active molecules, the peculiarity of which above all others is, hence, that they possess just this critical energy. The critical energy is assumed — in a given system — to be independent of the temperature, and it is assumed that all active molecules actually react. By consideration of the influence of the temperature upon the reaction-velocity MARCELIN has succeeded in finding the difference between the critical energy and the mean-energy of the molecules, hence, the average accretion of energy which will be necessary to render all inactive molecules able to react. This "relative critical energy", E_k, is, according to MARCELIN, for one gr. molecule equal to:

$$E_k = RT^2 \frac{d \ln k}{dT}$$

where R is the gas constant, T the absolute temperature and k the velocity-constant for the said process.

If now the process be light-sensitive it will be of considerable interest to compare the necessary accretion of energy E_k with the amount of the absorbed light-energy which is required in order to produce a certain reaction-velocity. I have executed this comparison partly for my own measurements of the oxidation of the hydrogen iodide, and partly in a couple of cases occurring in the literature which offered the necessary data. In the tables are besides denoted the amounts of energy, $nNh\nu$, which would be necessary if the primary process consisted in a split off of electrons or in a moving ot these from one stationary state to another.

I. Own Measurements.

The Oxidation of the Hydrogeniodide.

The temperature-coefficient for the dark-process according to Plotnikow[1] is 2,07[2]. Hence:

$$E_k = 83{,}15 \cdot 10^6 \frac{293 \cdot 303}{10} ln\ 2{,}07 = 53{,}7 \cdot 10^{10}\ \text{Erg.}$$

Table 1.

Wave-length	transformed mol. per sec. in 5 mm³ = n	nE_k	absorbed energy	$nNh\nu$
	Experiments with passage of air.			
436 μμ	$0{,}75 \cdot 10^{-11}$	4,0 Erg.	5,9 Erg.	21,0 Erg.
—	1,00 »	5,4 »	6,5 »	28,0 »
405	0,59 »	3,2 »	2,9 »	17,7 »
—	0,50 »	2,7 »	2,3 »	15,0 »
366	0,68 »	3,7 »	5,5 »	22,6 »
—	0,52 »	2,8 »	3,5 »	17,3 »

[1] Zeitschr. f. physikal. Chemie 64, 215 (1908).

[2] In an earlier paper (Zeitschr. f. physikal. Chemie 58, 214 (1907)) the same author had determined the temperature-coefficient at 2,86, which would render all of the E_k-values in the table somewhat greater, hence render them still more comformable.

Wave-length	transformed mol. per sec. in 10 mm³ = n	nE_k	absorbed energy	$nNh\nu$
313 μμ	$1{,}33 \cdot 10^{-11}$	7,2 Erg.	3,1 Erg.	51,8 Erg.
—	1,50 »	8,1 »	3,5 »	58,4 »
	Experiments with passage of oxygen.			
	transformed mol. per sec. in 5 mm³ = n			
366	$150 \cdot 10^{-11}$	8,1 »	3,7 »	50,0 »

II. Other Measurements.

The Decomposition of the Hydrogen iodide

The temperature-coefficient is according to Bodenstein[1]:

$$lnk = \div \frac{21922{,}5}{T} \div 14{,}468\, lnT + 0{,}023055\, T \div 104{,}185.$$

Photochemical decomposition and measurement of energy: Warburg[2]. The calculation is applied to experiment No. 19, where the transformation is greatest:

Wavelength	n	nE_k	E_{abs}	$nNh\nu$
282 μμ	$8{,}25 \cdot 10^{-6}$	$1{,}36 \cdot 10^{7}$	$1{,}64 \cdot 10^{7}$	$3{,}45 \cdot 10^{7}$ Erg.

The Decomposition of the Peroxide.

The temperature-coefficient for the platinum catalysis according to Bredig and M. v. Berneck[3]

$$ln \frac{k_1}{k_2} = 5899 \frac{T_1 - T_2}{T_1 T_2}$$

Photochemical decomposition and measurement of energy: Henri and Wurmser[4]

Wavelength	n	nE_k	E_{abs}	$nNh\nu$
280 μμ	1	$0{,}82 \cdot 10^{-12}$	$1{,}54 \cdot 10^{-12}$	$6{,}7 \cdot 10^{-12}$ Erg.

[1] Zeitschr. f. physikal. Chemie 29, 306 (1899).
[2] Sitzungsber. d. Berliner-Akad. 1918, 300.
[3] Zeitschr. f. physikal. Chemie 31, 329 (1899).
[4] Compt. rend. 157, 126 (1913).

The temperature-coefficient will undoubtedly be far greater for the peroxide-decomposition itself (without platinum) which — as known — only has a measurable velocity at high temperature, than for the platinum catalysis[1]. Putting $nE_k = E_{abs} = 1{,}54 \cdot 10^{-12}$, we find the temperature-coeficient equal to 3.52, (instead of 1,7 for the platinum catalysis), which is a far more probable value.

Even if the material is, for the present, very limited[2], all the evidence seems to suggest that the relative critical energy is equal to the absorbed energy, that the latter consequently is completely utilised, as it brings all of the absorbing molecules into an active state, i. e. a state capable of reacting. The deviations which are found, and which almost all tend towards the relatively critical energy being a little smaller than the absorbed, are explained by defective stirring. As far as can be seen from the description of experiments, stirring was not employed at all in Warburg's experiments, beyond the movement produced by slow passing-through of hydrogeniodide, and, the absorption being strong (70 per cent), and the bundle of rays only having been $^1/_6$ of the transverse section of the cell, the radiation cannot have been capable of exerting its full effect. Warburg's measurements show also that nE_k decreases with decreasing wave-length, i. e. with increasing absorption.

Quite the same thing holds also for Vranek's measurements of the decomposition of the potassium kobaltioxalate.

[1] Nernst (Zeitschr. f. Elektroch. 11, 710 (1905)) has for the decomposition of the pure peroxide found the temperature-coefficient 1,32 between 270 and 336°. But it is possible, partly that the process also under these circumstances is catalytic (the effect of the walls), and partly we cannot compute from this the magnitude of the coefficient at ordinary temperature.

[2] More accurate experiments are being prepared.

Table 2.

n	E_k	$n \cdot E_k$	Wavelength	E_{abs}	Extinction
0,0002	$1{,}12 \cdot 10^{12}$	$2{,}24 \cdot 10^{8}$	436 μ	$7{,}6 \cdot 10^{8}$ Erg.	180,4
»	»	»	405 μ	5,8 »	163,5
»	»	»	366 μ	4,3 »	58,5

For my experiments with the oxidation of the hydrogen-iodide the same holds good. Despite a very vigorous stirring I am not certain of having obtained the complete utilisation of the absorbed energy, (the absorption was in every place complete), the process occurring very rapidly. If we, f. inst. by making the distance to the lamp greater, were to decrease the velocity and thereby make the stirring more effective, the energy-measurement would simultaneously become so incorrect that hardly anything would be gained by this.

Other apparent deviations from the found regularity will be dealt with in a later section.

We may now achieve a corroboration of the correctness of the found law from a series of other measurements, not computable in the same way as those above, the temperature-coefficient not being known.

Taking it for granted that the temperature-coefficient $k_{t+10} : k_t$, which for processes with conveniently measurable velocity as a rule lies between 2 and 4, may rise to 10, when the process is considered at a far lower temperature than that at which its velocity is measurable[1], we may from the found »law of critical efficiency« compute the limits for the "specific photochemical effect", as it is defined by WARBURG. This "specific photochemical effect" d

[1] For ferric oxalate, the decomposition of which is not measurable until above 40°, we compute from LEMOINE's measurements (Ann. chim. phys. (6), 30, 289 (1893)) for 15—25° a temperature-coefficient of 8.5. For the decomposition of hydrogen chloride VAN'T HOFF computes 17.2 (Vorles. üb. theoret. u. physik. Chemie I, 231).

is the transformed amount of substance per calorie of the energy absorbed.

Hence:

$$\Phi = \frac{n}{E_{abs}}.$$

As, according to the law of critical efficiency, $nE_k = E_{abs}$,

$$\Phi = \frac{1}{E_k}.$$

From Marcelin's formula is computed, for the temperature-coefficient 2, $\Phi = 8{,}1 \cdot 10^{-5}$, for the temperature-coefficient 10, $\Phi = 2{,}4 \cdot 10^{-5}$.

The following examples are found in the literature:

Table 3.

Process	Wavelength	$\Phi \cdot 10^5$	Examin. by:
$NH_3 \longrightarrow N + 3\,H$	209 μμ	0,172	Warburg[1]
$O_2 \longrightarrow O_3$	253 »	0,98	»
»	209 »	1,35	»
$O_3 \longrightarrow O_2$	287 »	2,96—6,70	»
»	253 »	0,25—7,08	»
»	209 »	6,58—8,09	»
$HBr \longrightarrow H + Br$	253 »	1,79	»
»	209 »	1,53	»
$HJ \longrightarrow H + J$	282 »	2,09	»
»	253 »	1.85	»
»	209 »	1,44	»
Nitrobenzaldehyde → Nitrosobenzoicacid	405 »	ca. 2	Kummerer[2]
»	366 »	ca. 4	»

That many of these figures are smaller than the denoted lowest limit, cannot give rise to any astonishment. The reason may partly be the above, that a complete utilisation of the absorbed energy can only be expected when working in thin layers and with a stirring so vigorous, that the velocity of the light-process is very

[1] Sitzungsber. d. Berliner-Akad. 1911—18.

[2] Inang.-Dissert. Berlin 1914.

slight when compared with the equalisation of the concentration-difference produced by the stirring, and partly the deviations may arise from the fact, that the temperature-coefficient of the process after all is higher than the highest assumed limit.

Above all it must be borne in mind that nothing whatever can be stated beforehand in any other way about the magnitude of the specific photochemical effect, and that the values might consequently be imagined to be spread over a very large interval, while they actually lie within fairly narrow limits, a circumstance which may be naturally explained by the fact that the temperature-coefficients, too, only vary comparatively little.

From WARBURG's measurements we may deduce yet another corroboration of the law of critical efficiency. The reversible darkness-process:

$$A \rightleftarrows B$$

is supposed to be displaced to the right by heating, which means that the process $A \rightarrow B$ has a greater temperature-coefficient than the opposite process. It follows, consequently, that the process $A \longrightarrow B$ must also have a greater relative critical energy, hence, a smaller specific photochemical effect than the process $B \rightarrow A$. By consideration of table 3 it is seen in accordance herewith that Φ for all wavelengths is smaller for the process $O_2 \longrightarrow O_3$ than for the reverse, which agrees with the fact that the equilibrium by heating is displaced in favour of the formation of ozone.

II. Consequences of the Law of Critical Efficiency.

If the here found law of critical efficiency is corroborated on closer examination, it will firstly yield a highly

welcome means for independent determination of the relative critical energy, a quantity which will undoubtedly in future play a very great rôle in the further development of general chemistry.

Secondly, it evidently means a great step towards a quantitative control of the photochemical processes, as, for the first time, the possibility is here present that — independent of all hypotheses of the special mechanism of the primary reaction — we may beforehand calculate the light-sensitiveness of the process, if only the temperature-coefficient of the corresponding darkness-reaction be known.

Furthermore the law of critical efficiency throws light on various photochemical relations which have hitherto not been explainable, thus f. inst. the smallness of the temperature-coefficients. The temperature-coefficient of the dark-processes may in its dependency on the temperature most frequently be expressed by an equation such as:

$$lnk = \div \frac{A}{T} \div BlnT + CT + D,$$

where B, C and D are far smaller than A. By differentiation we obtain:

$$\frac{dlnk}{dT} = \frac{A}{T^2} \div \frac{B}{T} + C.$$

Now:

$$E_k = RT^2 \frac{dlnk}{dT}.$$

If in the first approximation we put B and C equal to zero, we obtain for two different temperatures:

$$\frac{E_k'}{E_k} = \frac{\Phi}{\Phi'} = \frac{T^2 \cdot T_1^2}{T_1^2 \cdot T^2} = 1.$$

$\Phi' : \Phi$ is the temperature-coefficient of the light-process, which will accordingly be 1 or — when all the terms in the formula for *lnk* are included — at any rate very close to 1. Now it has long been known[1] that temperature-coefficients for photochemical processes are generally far lower than for the dark-processes, and in many cases approach very closely to 1.

We also, however, know several light-processes, with temperature-coefficients around 1,2 and 1,4, and even a very few between 2 and 4, which would require a quite absurdly big alteration of the temperature-coefficient of the dark-process with the temperature. Now it must be remembered that all of these temperature-coefficients are computed for the same incident amount of light, not for the same absorbed amount of light. But in the above argument I have taken it for granted that E_{abs} was uniform in the two considered temperatures which need not be the case at all. If this is not the case, we have at two temperatures:

$$n_1 E_k = E_{abs}$$
$$n_2 E'_k = E'_{abs}.$$

The temperature-coefficient $n_1 : n_2$ is then

$$\frac{n_1}{n_2} = \frac{E_{abs} \cdot E'_k}{E'_{abs} \cdot E_k}$$

In order to reach the temperature-coefficients usually given, we must consequently divide the fraction $E'_k : E_k$ by the ratio between the amounts of light which are absorbed at the two temperatures in question, and the more this ratio deviates from 1 the more the temperature-coefficient will also deviate from 1.

[1] E. Goldberg, Zeitschr. f. wiss. Photogr. 4, 103 (1906).

Another peculiar circumstance with regard to the photochemical temperature-coefficient is also explained by the law of critical efficiency. In several recent works Padoa and his co-workers[1] have shown for a series of photochemical processes, that the temperature-coefficient for a single process most frequently increases with the wave-length. The examined processes are:

Table 4.

	Temperature-coefficient			
Light-sensitive system	Green	Blue	Violet	Ultra-violet
Chlorine-hydrogen mixture	1,50	1,31	1,21	1,17
Eder's liquid	1,75	1,21	—	1,05
Benzaldehydephenylhydrazone	—	1,08	1,04	—
Triphenylfulgide (coloration).....	1,67	1,29	—	—
Salicyliden-β-naphtylamine	1,80	1,45	1,39	—
	Red	Yellow		
Triphenylfulgide (de-coloration) ...	1,08	0,89	—	—

On the other hand 1.05 was found independent of the colour for photographic plates.

Now it is a fact that, by heating the absorption-curve is most frequently displaced towards the long-waved end of the spectrum. This must involve that the relation $E_{abs} : E'_{abs}$ must increase the more we move from the absorption-maximum towards the long-waved end of the spectrum, whereas it must decrease when we move the opposite way. As according to the above, it is materially this circumstance which determines the magnitude of the photochemical temperature-coefficient, it is intelligible that it varies with the wave-length and just in the way shown by the observations. For chlorine-hydrogen mixture, Eder's liquid, benzaldehydephenylhydrazone and salicyliden-β-nafthylamine the absorption-maximum everywhere lies in

[1] Atti d. Reale Accad. d. Lincei (5) 22 II 500 (1913). 24 I 828 (1915). 24 II 97 (1915). 25 I 808 (1916). 25 II 168, 215 (1916).

ultraviolet. The temperature-coefficient must accordingly either keep independent of the wave-length — if the absorption is independent of the temperature — or rise towards green. Triphenylfulgide has its absorption-maximum in blue-violet, and must consequently behave in the same manner. The (phototropic) triphenylfulgide, coloured by the insolation is, on the contrary, blue, consequently has absorption-maximum in orange. Towards red on the long-waved side of the maximum, the temperature-coefficient must then rise above 1, towards yellow, on the short-waved side of the maximum it must, on the contrary, fall below 1, as is also shown by the observations.

Also the possibility of chemi-luminescence may be elucidated from the law of critical efficiency. According to MARCELIN[1] the critical energy is uniform for the two opposite processes in a reversible process. If the mean-energies for the two processes are E_1 and E_2, the relative critical energies E_k and E'_k, then consequently:

$$E_1 + E_{k'} = E_2 + E'_k.$$

The energy yielded after the electron has "passed the top", E'_k, must presumably be yielded in quanta, the frequency of which is computable according to the formula:

$$E'_k = Nhv$$

where N is the figure of AVOGADRO, h PLANCK's constant and v the frequency. As a resource in the calculation of E'_k in those cases where it cannot be directly measured (or is not known), we make use of the circumstance, that the difference between E_k and E'_k is equal to the reaction-heat in the said process. In this way I have in the table

[1] Ann. de physique 3, 173 (1914).

below computed E_k firstly for the process $O_3 \rightarrow O_2$ starting from the specific photochemical effect for the opposite process, which is assumed to be correct. Hereby we obtain the value $2{,}7 \cdot 10^{-5}$ for Φ, which is more probable than Warburg's experimentally found value $1{,}35 \cdot 10^{-5}$, which, as earlier mentioned, is undoubtedly too low on account of the lack of stirring. Furthermore for several other processes where E_k and the reaction-heat is known. The reaction-heat Q as well as the other energy-quantities are expressed in erg.

Table 5.

Process	$E_k \cdot 10^{-10}$	$Q \cdot 10^{-10}$	$E'_k \cdot 10^{-10}$	$\upsilon \cdot 10^{-14}$	Wavelength
$O_2 \rightarrow O_3$	155	$-$ 95	60	1,52	1970 μμ
$O_3 \rightarrow O_2$	60	$+$ 95	155	3,93	764 »
$2\,HJ + O \rightarrow J_2 + H_2O$	53,7	$+$ 30,8	84,5	2,15	1390 »
$HJ \rightarrow H + J$	200	$+$ 12,6	212,6	5,4	557 »
$H + J \rightarrow HJ$	212,6	$-$ 12,6	200	5,1	592 »
$HBr \rightarrow H + Br$	234	$-$ 50,6	183,4	4,7	642 »
$H + Br \rightarrow HBr$	183,4	$+$ 50,6	234	6,0	503 »
$NH_3 \rightarrow N + 3\,H$	2430	$-$ 50	2380	60	50 »
$N + 3\,H \rightarrow NH_3$	2380	$+$ 50	2430	62	48 »

This calculation is of course only of interest on the assumption that the measured reaction-heat correspond to the simple reaction scheme which is given in the table, E'_k thus being rendered without any break, without division processes. On this assumption the table shows that the decomposition and formation both of the hydrobromic acid and the hydriodic acid, as far as these processes could be made to occur rapidly in darkness at ordinary temperature, must have a possibility of luminescing in green-orange, and the transformation ozone → oxygen in red. The transformation oxygen → ozone and the oxidation of the hydriodic acid, on the contrary, emits

ultra-red radiation, and the decomposition and formation of ammonia emits ultra-violet radiation in the not so easily accessible domain below 200 $\mu\mu$. About all these processes it holds true, however, that at present we do not know any means to increase their reaction-velocity at ordinary temperature to such a degree that there can be any possibility of demonstrating or measuring this secondary radiation. Several of them even represent displacements of equilibria, which cannot be made at all without external supply of energy. The calculation is only carried through in order to show, how a luminescent system (f. inst. pyrogallol in alcaline liquid + an oxidizing substance or another of the numerous chemi-luminescent processes) should be investigable, if we start from the view here maintained. Furthermore the result has the interest of showing that the found frequences after all lie within the region with which we can on the whole reckon when the question is of radiation from atoms and electrons.

III. Deviations from the Law of Critical Efficiency.

A regularity like the photochemical law of critical efficiency, which is at least made probable by the experiments and considerations which have been described in the foregoing sections, must — according to its nature — be assumed to be generally valid, in the same way as, f. inst. the law which describes the distribution of a substance between two phases. But just as the last named law in many cases seems to be invalid by the fact that the substance in one phase may be found in several different conditions or compounds, while the law only reckons with the same kind of molecules in both phases, thus it

might also be expected beforehand that many apparent deviations from the photochemical law of critical efficiency may be found. In the following I will go through some of the causes of the deviations which can be recognized for the present. It will be natural to divide these into positive and negative deviations.

The positive deviations, by which the measured light-sensitiveness is greater than the computed, may as a rule be caused by the primary formation of a catalyst. We can here as little as in the darkness-catalyses indicate anything beforehand as to how much such a catalyst is able to accelerate the process or how great the "induction-factor" is, in so far as the catalyst itself is consumed in the course of the process. Consequently no limit can be given for the total light sensitiveness, which can be obtained, when a very active catalyst is formed in the primary process.

A typical example of this kind we have in the catalysis of the peroxide with potassium-ferro-cyanide[1]. Under the circumstances applied in this investigation (left for 30 minutes) the induction-factor has the following values:

Table 6.

$c_{H_2O_2}$	λ	$E_k \cdot 10^{-5}$	$E_{abs} \cdot 10^{-4}$	$E_k : E_{abs}$
0,04-m	405	3,3	0,73	45
»	»	3,3	0,70	48
»	366	39	9,8	40
»	»	54	10,0	54
»	»	72	10,7	67
»	»	113	21,3	53
				51
0,02-m	»	64	17,2	37
»	»	152	38,6	39
				38

[1] Kgl. Danske Vid. Selsk. Math.-fys. Medd. II. 1.

The light-catalyst here being exceedingly constant, we should obtain a far higher induction-factor by leaving the mixture in the dark for a longer period. In experiment 13 (loc. cit.) we have, f. inst. by leaving for $2^1/_2$ hours, the induction-factor 270, when the decomposition in the dark is subtracted.

Quite an analogous case is offered by the photographic plate. The latent image, at the formation of which the law of critical efficiency might undoubtedly be applicable, here represents the light catalyst. It is at least as constant as the catalyst formed by exposure of potassium-ferrocyanide, and the total process has (by the development) a very high induction-factor. How great it is, is not known, the few quantitative determinations of the light-sensitiveness at hand only referring to the incident and not to the absorbed light.

To the same group probably also belongs the reaction in chlorine-hydrogen mixture which, from its discovery, has been reckoned as one of the most sensitive of all known photolyses. Bodenstein[1] computes the quantum sensitiveness at 10^6, but an actual determination of the sensitiveness has never been carried out.

While the positive deviations from the law of critical efficiency can no doubt mainly be explained from this point of view[2] the negative deviations may originate from various sources. We have previously (v. p. 16) mentioned a purely technical cause, the lack of stirring. A frequently occurring case is, that besides the actual light-sensitive substance another, very absorbing, substance is found. The

[1] Zeitschr. f. physikal. Chemie 85. 346 (1913).

[2] Provided, that it is quite the same process, which is measured in darkness and in light.

measured absorption will then only partly be of any good to the actual sensitive substance. As an example may be named the oxidation of the quinine with chromic acid, where Luther and Shannon Forbes[1] have succeeded in finding the part of absorbed light-energy which belong to the quinine, which is here the sensitive substance. Unfortunately the two investigators have not measured this absorption in absolute measure, and the dark-temperature-coefficient of the process is not known either, as far as we know, for which reason the calculation cannot be carried through.

If the non-sensitive, absorbing substance be formed by the sensitive substance, the concentration of the latter is decreased to a frequently unknown degree. The sensitiveness calculated for the total amount of this substance will then always be far smaller than the actual sensitiveness.

While these examples deal with injurious absorption of the incident, active light, it might also be imagined, that the secondary radiation, to which — according to a hypothesis advanced in another paper — the absorbed radiation in many cases is transformed, was absorbed by other substances than that upon which it was actually to act. Deviations, which perhaps may be explained in this way, is found in Warburg's works.

While the specific photochemical effect of the decomposition of the hydriodic acid and the hydrobromic acid is independent of the concentration of these substances, hence, is constant, Warburg[2] finds highly peculiar circumstances in the photochemical des-ozonisation (Table 7).

The specific photochemical effect which, according to

[1] Journ. amer. chem. Soc. **31**, 770 (1909).
[2] Sitzungsber. der Berliner Akad. **1913**, 644.

Table 7.

Ozone, solved in:	Wave-length	Ozone-Concentration	Absorption of the Ozone	$\Phi \cdot 10^5$
Oxygen	253	9,29 —0,75	1 —0,94	3,13 —0,44
»	»	0,47 —0,18	0,752—0,288	0,253
Nitrogen	»	2,56 —0,65	0,98 —0,88	2,37 —0,99
»	»	0,65 —0,167	0,877—0,267	0,97
Helium	»	0,429—0,361	0,804—0,710	1,60 —1,44
Oxygen (moist.)	»	8,49 —0,91	1	7,08 —1,84
»	»	10,50 —0,80	1 —0,86	4,23 —1,49
»	209	8,49 —5,66	0,857—0,733	8,09 —6,58
»	287	8,15 —6,27	0,96 —0,92	6,70 —2,96

the law of critical efficiency should be constant, thus varies very much, both with the concentration of the ozone and the nature of the solvent.

Let us, f. inst. consider the mixture of ozone and oxygen. By exposure to 253 μμ only the ozone absorbs to a mentionable degree so that nothing could be lost in this way. Now WEIGERT[1] has made it probable, that a light catalyst is formed by the photochemical desozonisation. According to the radiation-hypothesis this catalyst acts by transforming the absorbed radiation quantitatively to a far shorter wave-length, which may also be absorbed by the ozone. Just as iodine-ion and tri-iodine-ion, the leuco-compounds and the corresponding colour-substances possess a common absorption-region in ultra-violet[2], we may be allowed to consider it probable that the oxygen, too, has the same absorption-region in furthest ultra-violet, as its oxidation product ozone. This means that the oxygen consumes part of the secondary radiation which was to be employed for the desozonisation.

The experiments now show, firstly, that the specific

[1] Zeitschr. f. physikal. Chemie **80**, 78 (1912).

[2] Kgl. Danske Vid. Selsk. Math.-fys. Medd. II. 2.

photochemical effect for the same ozone concentration is higher when ozone is mixed with nitrogen, than when mixed with oxygen, and still higher, when mixed with helium. According to our knowledge of the absorption in furthest ultra-violet[1] of these three gases, we seem justified in supposing that oxygen has the greatest, helium the smallest extinction of the three. The secondary radiation of the ozone might therefore be best utilised for decomposition of ozone when the surrounding molecules consist of helium, not so well when the ozone is diluted with nitrogen and least when mixed with oxygen, as also shown by the experiments.

In very small ozone-concentrations Warburg finds for the mixture with oxygen and nitrogen, that the specific photochemical effect becomes fairly constant, even if the ozone-concentration, and with it the absorption, vary. This constancy arises by compensation of two inverse influences.

According to Warburg's statement, the radiation passed through a 0,95 cm. high and 0,43 wide split into the circular cell, the diameter of which was 1,7 cm and the thickness 1,9 cm. and spread somewhat there. When no stirring is employed only a fraction of the contents of the cell would consequently be able to be exposed during the 4 minutes each experiment lasted. But the measured loss of ozone corresponds to all of the contents of the cell, hence, is smaller than the loss of ozone in the exposed part of the cell. The smaller the ozone-concentration is the greater is the exposed part of the cell — on account of the decreasing absorption — the greater, consequently

[1] Lyman: "The Spectroscopy of the Extreme Ultra-Violet", p. 63, 69 (1914).

also the measured specific photochemical effect in proportion to the correct effect. This augmentation of Φ with decreasing ozone-concentration is, however, counteracted by the loss described above by the absorption of the solvent of the secondary radiation, and it is at any rate possible that the combination of these two opposite influences can produce the approximate constancy, which has been found by WARBURG.

Suppose v to be the volume of the cell, v_1 the highly exposed, and v_2 the slightly exposed and dark portion of the cell. The total ozone-concentration is c, the loss of concentration in the exposed part c_1, and the measured loss of concentration c_2. We then have:

$$cv_2 + (c - c_1)v_1 = (c - c_2)v,$$

from which:

$$c_2 = c_1 \frac{v_1}{v}, \text{ since } v = v_1 + v_2.$$

As mentioned v_1 is, however, not constant, but increases with decreasing absorption. The constancy of Φ occurring when the absorption is abt. 0,9, the highly exposed volume may suitably be put proportional to the thickness of layer x, which at different concentrations just brings about this absorption. If α be the extinction-coefficient we have:

$$1 - 10^{-\alpha c x} = 0{,}9, \text{ then } x = \frac{1}{\alpha c} \text{ and } v_1 = \frac{k_1}{\alpha c}.$$

The specific photochemical effect which was actually to be measured, is now pro primo the transformed amount of substance per absorbed unit of energy:

$$\frac{c_1}{I_0 (1 - 10^{-\alpha c_1})},$$

where I_0 is the intensity of the incident radiation. By the made measurement c_2 is, however, introduced instead of c_1. Pro secundo the specific photochemical effect is furthermore decreased by the solvent's absorption of the secondary radiation, the measured specific photochemical effect Φ_2 thus being:

$$\Phi_2 = \frac{c_2}{I_0 (1 \div 10^{\div \alpha c})} \cdot \frac{\alpha_1 c}{(\alpha_1 c + \alpha_2 (100 \div c))},$$

where α_1 and α_2 are the extinction-coefficients of the ozone and oxygen respectively for the secondary radiation.

In small ozone-concentrations the absorption — as shown by Warburg — is very near being proportional to the concentration, therefore the quantity $1 \div 10^{\div \alpha c}$ may be replared by αc. Secondly the term $\alpha_1 c$ becomes very small in comparison with $\alpha_2 (100 \div c)$ and c very small in comparison with 100. By introducing the values for c_2 and v_1 the expression will consequently be:

$$\Phi_2 = \frac{c_1 k_1 \alpha_1 c}{I_0 \alpha c v \alpha c \alpha_2 100} = K_1 \frac{c_1}{c}.$$

The specific photochemical effect, which was actually to be measured, is — when all disturbing influences are imagined to be removed —:

$$\Phi_1 = \frac{c_1}{I_0 (1 \div 10^{\div \alpha c})},$$

or for small concentrations:

$$\Phi_1 = \frac{c_1}{I_0 \alpha c} = K_2 \frac{c_1}{c}.$$

According to the law of critical efficiency this quantity only depends upon the temperature-coefficient of the darkness-process, hence it is constant, and Φ_2 being propor-

tional to the ideal Φ_1, Φ_2 must be constant, as found by WARBURG, and possess the value:

$$\Phi_2 = \frac{k_1 \Phi_1 \alpha_1}{\nu \alpha \alpha_2 100}.$$

The difference between the measured specific photochemical effects in the mixtures with oxygen, nitrogen and helium then only depend upon the different value of α_2. If for oxygen (see below) it be put equal to α_1, then it will be for:

$$\text{Nitrogen: } \alpha_2 = \alpha_1 : \frac{0{,}97}{0{,}213} = \frac{\alpha_1}{3{,}8}$$

$$\text{Helium: } \alpha_2 = \alpha_1 : \frac{1{,}52}{0{,}253} = \frac{\alpha_1}{6{,}0}$$

by which helium absorbs the secondary radiation 6 times, nitrogen 3,8 times less than oxygen (and ozone).

Also in somewhat greater concentrations Φ_2 must be fairly constant as long as α is so, and the diffusion does not play any part. If we put:

$$\frac{c_1}{I_0 (1 \div 10^{\div \alpha c})} = \Phi_1$$

in the expression of Φ_2:

$$\Phi_2 = \frac{c_1 k_1}{\nu \alpha c I_0 (1 \div 10^{\div \alpha c})} \frac{\alpha_1 c}{(\alpha_1 c + \alpha_2 (100 \div c))}$$

we have:

$$\Phi_2 = \frac{k_1 \Phi_1 \alpha_1}{\nu \alpha (\alpha_1 c + \alpha_2 (100 \div c))}.$$

As long as c only amounts to a few per cent, and α_1 is not very large when compared with α_2, we shall always be able to put the expression $\alpha_1 c + \alpha_2 (100 \div c)$ equal to 100, the error only amounting, perhaps, to a few per cent, which falls quite within the error-limit in WARBURG's experiments.

The formula will accordingly be:

$$\Phi_2 = \frac{k_2}{\alpha}.$$

The experiments, however, exhibit both for the mixtures with oxygen and with nitrogen a fairly high increase of Φ_2, with the ozone-concentration. But the explanation is near at hand. Table 8 contains the data in question and besides those extinction-coefficients, α, which are computed from c and the relative absorption A.

Table 8.

c	A	α	$\Phi_2 \cdot 10^5$	$\alpha\Phi_2 \cdot 10^5$
		Ozone and Oxygen		
9,29	1	—	3,13	—
2,92	0,95	0,45	1,48	0,67
1,85	0,98	0,92	0,68	0,63
0,75	0,94	1,63	0,44	0,72
0,47—0,18	0,75—0,29	1,60	0,253	0,41
		Ozone and Nitrogen		
2,56	0,98	0,66	2,37	1,57
1,41	0,93	0,82	1,70	1,40
0,65	0,88	1,42	0,99	1,40
0,46—0,17	0,74—0,27	1,60	0,97	1,55

The table shows the astonishing circumstance that the extinction-coefficient decreases very much with increasing ozone-concentration. Such a deviation from Beer's law is not unknown in the case of gases. Warburg[1] has for oxygen found the inverse relation, that the extinction-coefficient increases more than the oxygen-pressure. If we take regard to this variation of the ozone, the table shows that the product $\alpha\Phi_2$ is constant, as demanded by the formula. With regard to the mixture with nitrogen this constancy extends over all the region, whereas at the mixture with oxygen the most diluted mixtures deviate

[1] Sitzungsber. d. Berliner-Akad. 1915, 230.

pretty much from the others, without the reason for this being discoverable from the description of the experiments.

If we rise to still greater concentrations there are two cases to consider; the expression $\alpha_1 c + \alpha_2 (100 - c)$ cannot anymore be put constant, and the diffusion here begins to operate. WARBURG's measurements of greater ozone-concentrations now only comprise mixtures with oxygen, where there might be reason to assume that α_1 and α_2 are fairly uniform. COEHN and STUCKHARDT[1], for the analogous pair of substances iodine-hydrogeniodide have found that the two substances not only possess approximately the same absorption-region for ultra-violet, but also nearly the same extinction. In the same way it must be allowable to assume in the first approximation, that the oxygen has the same absorption-region and extinction in furthest ultra-violet, as its oxydation-product, the ozone. Thereby the fraction will be:

$$\frac{\alpha_1 c}{\alpha_1 c + \alpha_2 (100 - c)} = \frac{c}{100}.$$

The velocity of diffusion can be put proportional to the difference in the concentration between the exposed and the non-exposed part of the cell, hence to c_1. For this reason no other alteration in the formula for Φ_2 takes place than the introduction of a new constant. After contraction the found specific photochemical effect will be:

$$\Phi_2 = \frac{k_2}{\alpha}.$$

As shown above α decreases with increasing ozone-concentration, whereby the fairly large rise in Φ found by WARBURG in other experimental series is explained.

[1] Zeitschr. f. physikal. Chemie **91**, 722 (1916).

The values of α, denoted in table 8, are, however, too incorrect for us to determine their dependency on the concentration with any certainty, nor is it for this reason possible to put forth any certain relation between Φ and c.

Results of this work.

The classification of the photochemical processes into work-accumulating and work-yielding processes does not agree with our present knowledge of the nature of the photolyses, but is essentially of technical interest.

On comparing the energy-absorption, necessary for the oxidation of a certain amount of hydriodic acid, with the energy which according to Marcelin's calculation is necessary in order to render the same quantity of hydriodic-acid molecules able to react (active), it appears that these two quantities of energy are identical. The absorbed energy is, hence, quantitatively utilised in order to make the absorbing molecules able to react. This law is verified in two other processes.

This law of critical efficiency explains the peculiar circumstance that the "specific photochemical effect" (Warburg) for a series of very different processes is of the same order of magnitude.

The law of critical efficiency renders it possible in many cases to calculate the light-sensitiveness of a process beforehand, if only its darkness-temperature-coefficient be known.

The law of critical efficiency gives a natural explanation of the circumstance that the temperature-coefficients of

the photochemical processes are generally very small and sometimes increase with the wave-length.

The law of critical efficiency elucidates the possibility of chemi-luminescence.

The law of critical efficiency gives a probable explanation of the empirically found rule, that the less sensitive a process is to a certain wave-length, the more easily accessible it is to the optical sensitising for this wave-length, and vice versa.

The deviations occurring from the law of critical efficiency may partly be explained by experimental errors, as insufficient stirring or lack of knowledge of the absorption of the actual sensitive substance, partly by the formation of a light-catalyst, and finally by the assumption, which is supported by the observations on zinc oxide (which will be published later on) and the oxidation of the hydriodic acid, that the absorbed radiation in many cases is quantitatively transformed to radiation of a far smaller wave-length. The deviations from the law of critical efficiency are then due to the fact that this secondary radiation is not only absorbed by the substance, the transformation of which is to be effected hereby, but also by other substances.

Royal Technical College of Copenhagen,
May 1919.

Færdig fra Trykkeriet den 10. Februar 1920.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. II, 4.

SUR
L'ORIGINE DE L'ALGÈBRE

PAR

H. G. ZEUTHEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL.
BIANCO LUNOS BOGTRYKKERI
1919

I.

De nos jours, les recherches algébriques se font au moyen d'un système de symboles représentant, par lettres, des quantités quelconques et, par signes, les opérations auxquelles on soumet les quantités. L'histoire du développement de ce système et de l'extension successive du domaine auquel on l'a appliqué présente certainement beaucoup d'intérêt. L'emploi d'un tel système fut préparé par l'usage de certaines abréviations dans la langue courante. Déjà Aristote se sert de lettres pour représenter des quantités quelconques, Diophante fait usage de certaines abréviations pour la représentation, non seulement d'une quantité inconnue, mais aussi de ses puissances aux exposants $-6 \ldots . +6$ — comme on dirait de nos jours —, et il exprime l'addition par une juxtaposition, la soustraction par un signe particulier. L'emploi de signes se généralisa vers la fin du moyen âge. En systématisant l'usage croissant des différents symboles, Viète composa une *Logistica speciosa* capable d'exprimer les différents calculs exécutés sur des quantités quelconques conformément à des règles bien déterminées. Indirectement, Viète suppose pourtant encore que les quantités soient rationnelles; car ses signes d'opération représentent des calculs qui n'étaient définis que pour de telles quantités. Voilà pourquoi il ajoute à l'emploi de ses règles des démonstrations qu'il appelle

géométriques, et dont la généralité est une conséquence de celle de la théorie des proportions établie par Eudoxe. Descartes évite les renvois réitérés en donnant aux opérations exprimées par les symboles des définitions généralisées qui se rattachent directement à la théorie des proportions. Plus tard, on les a généralisées ultérieurement en appliquant les mêmes symboles à des quantités négatives et imaginaires et même à des opérations mathématiques d'un ordre plus élevé.

L'introduction successive et le développement historique des différents symboles mériterait bien de faire objet d'un exposé particulier.[1] Cependant, pour bien comprendre le besoin qui a conduit à la création de tels symboles et au développement ultérieur de leur emploi, il faut commencer par se rappeller la connaissance antérieure des opérations elles-mêmes pour lesquelles on essayait de créer un nouvel organe, et la généralité qu'on avait déjà su donner à ces opérations en les représentant d'une manière toute différente. L'introduction des nouveaux symboles se rattachait avant tout à la connaissance déjà acquise des équations du second degré. Par sa simplicité, la nouvelle représentation des équations et de leurs solutions en rendait la théorie plus accessible, l'application pratique plus facile. L'usage des mêmes symboles devait donc paraître tout indiqué à ceux qui se proposaient d'étendre la même théorie et d'en tirer de nouvelles conséquences en essayant d'appliquer des procédés semblables à des équations d'ordre supérieur ou d'en trouver de nouveaux qui y fussent plus applicables.

[1] Les nombreuses analyses qu'on doit à la plume de M. Enestrŏm portent à croire que, par ses connaissances spéciales des faits que doit contenir un tel exposé, il serait excellemment qualifié pour entreprendre cette tâche.

En même temps qu'on introduisait les nouveaux symboles, on gardait pourtant certaines dénominations, qui rappellent les moyens dont on s'était servi jusqu'alors pour représenter les opérations et obtenir les vérités qu'on avait connues et démontrées avant de penser à se servir d'un nouvel organe. L'usage algébrique d'appellations telles que carrés, cubes et les noms hypergéométriques de puissances supérieures montre que l'algèbre déjà connue qu'on allait symboliser d'une manière nouvelle s'était servie jusqu'alors de symboles géométriques, comme cela a lieu dans l'ancienne mathématique grecque, et ce que nous avons dit de Viète et de Descartes montre qu'en apprenant mieux la géométrie de l'antiquité on y avait encore recours pour donner à l'usage des nouveaux symboles la même précision et généralité qu'on trouvait et admirait chez les anciens auteurs.

C'est de l'algèbre ancienne, connue avant l'invention des symboles algébriques modernes, que je vais parler ici. C'est aussi à elle, mais nullement à sa première origine que remonte le mot algèbre. On sait que ce nom de notre science est emprunté au titre: *Al-dschebr w'al-muḳâbala* du premier traité d'algèbre arabe, celui d'al-Chwârizmî, et qu'il désigne les deux opérations, dont on devait faire usage pour donner à une équation la forme à laquelle s'appliquent les règles relatives à la solution d'une équation du second degré. Al-Chwârizmî ne s'occupe pas d'ailleurs de ces deux opérations, prescrites déjà par Diophante, et que leurs noms particuliers désignent comme ayant été familières aux géomètres arabes, car il ne trouve pas même nécessaire de les expliquer dans son livre. Par contre, il donne des règles précises pour la solution des différentes formes d'équations du second degré, qu'il illustre par des exemples numériques.

et il y joint des démonstrations des règles énoncées. Les équations et les règles sont exposées en paroles, et l'auteur ne fait pas même usage des abréviations de Diophante, mais les démonstrations sont faites, comme dans l'ancienne géométrie grecque, au moyen de figures contenant des rectangles et carrés qui représentent des produits de deux quantités différentes ou égales.

L'algèbre d'al-Chwârizmî, qu'il applique ensuite à la solution de questions auxquelles les lois de succession arabes donnaient lieu, ne diffère donc aucunement de celle que possédaient déjà les Grecs.[1] Néanmoins on a regardé longtemps son œuvre comme l'origine de l'algèbre, science qui lui doit son nom. Cela vient de ce que, au moyen-âge, l'algèbre de l'Occident était une continuation de celle des Arabes, dont du reste le développement ultérieur avait également été influencé par des suggestions dues à l'étude des mathématiciens grecs. Plus tard, les savants de la Renaissance voyaient bien que les auteurs grecs de l'antiquité étaient déjà en possession des procédés servant à résoudre les équations du second degré, et qu'ils savaient les employer dans leurs recherches géométriques; mais, dans la suite, et à mesure qu'on s'accoutumait à regarder l'algèbre comme inséparable de la symbolisation moderne, on inclinait à regarder les procédés, exposés sous forme géométrique par les anciens auteurs, comme des méthodes géométriques grâce auxquelles ces auteurs auraient bien su obtenir dans la géométrie la même chose que nous offre à présent un emprunt à l'algèbre, mais dont les mêmes auteurs auraient

[1] Qu'elle soit due à des traditions grecques contenues dans des écrits que nous possédons encore ou qui sont perdus, c'est ce que M. Cantor démontre d'une manière plus détaillée aux pp. 723—726 de ses Vorlesungen über die Geschichte der Mathematik I (3e édition).

— pensait on — ignoré l'application aux équations numériques.

Ce fut donc presque une découverte quand NESSELMANN constata (Die Algebra der Griechen, 1842) que l'emploi numérique d'équations du second degré était parfaitement familier à DIOPHANTE. Depuis, M. CANTOR a montré la même chose pour HERON, et PAUL TANNERY a observé que l'étude d'EUCLIDE, dans son livre X, sur l'irrationalité des quantités déterminées par des équations du second degré, montre qu'il visait aussi les équations numériques. EUCLIDE connaissait donc très bien cette application de procédés, que néanmoins dans ses Éléments il n'expose que sous forme géométrique et d'une manière très généralisée.

Cependant, cette succession des découvertes historiques relatives aux connaissances des Grecs a porté à attribuer un ordre semblable à leur propre acquisition de ces connaissances. On a cru que pour les Grecs eux-mêmes le développement de la théorie géométrique et les applications géométriques avaient été le but primordial, et que la solution d'équations numériques était pour eux une heureuse application dont on ne découvrit que plus tard la possibilité. Alors le développement aurait été l'inverse de ce que nous rencontrons ordinairement dans l'histoire des mathématiques. On commence presque toujours par inventer des procédés servant à résoudre quelques problèmes particuliers qui se présentent d'eux-mêmes; on en aperçoit de nouvelles applications, et on s'élève ainsi peu à peu à en reconnaître la portée et ensuite à donner un exposé bien établi de la méthode générale — ce n'est que la postérité qui croit voir dans le début de cette méthode la véritable source de la connaissance de toutes les vérités qu'on en peut déduire. Qu'un tel développement puisse s'étendre sur des siècles,

c'est ce que montre celui de la symbolisation moderne de l'algèbre. Utilisée d'abord dans les équations numériques, dont on avait appris la solution par l'intermédiaire des Arabes, et employée ensuite pendant la Renaissance à la solution de questions plus élevées que celles dont s'étaient occupés les anciens, ce n'est que dans la Géométrie de Descartes qu'elle s'élève à une généralité et une rigueur comparables à ce qu'on avait connu dans la représentation géométrique d'Euclide. Ce serait sans doute méconnaître l'état de la science caractérisé par ces propriétés que de croire que l'algèbre ancienne aurait pu naître sous une forme aussi élevée pour n'être qu'ensuite appliquée à des questions plus simples. L'excellence de la forme doit être au contraire le résultat d'un long développement commencé par des recherches plus ou moins fortuites et faites de points de vue moins abstraits; ces recherches antérieures ont dû ressembler à celles d'Al-Chwârizmî, beaucoup plus que ne font les propositions générales d'Euclide.

En réalité, la forme générale sous laquelle Euclide énonce et expose la solution des équations du second degré dans son livre VI, 27—29, est le résultat d'une transformation d'une algèbre plus ancienne, transformation qui devait accompagner la réforme générale qu'a subie la mathématique dans l'époque de Platon à Euclide; c'est du moins ce que j'ai tâché de montrer dans un mémoire sur cette réforme, publié en 1917[1]. J'y fais voir que dans la nouvelle géométrie scientifique on remplaçait, par la construction bien définie des figures déplacées, la simple transposition intuitive

[1] Hvorledes Mathematiken i Tiden fra Platon til Euklid blev rationel Vidénskab. Avec un résumé en français (Mémoires de l'Académie Royale des Sciences et des Lettres du Danemark, 8e série, Section des Sciences t. I nº 5 (1917)). Dans ce qui suit, je citerai ce travail comme mon »Mémoire sur la Réforme«.

de rectangles et de carrés qui avait servi déjà avant cette époque à symboliser les opérations algébriques de la même manière qu'elle le fait encore dans le livre d'Al-Chwârizmî, et qu'on généralisait ultérieurement les figures en substituant aux rectangles et aux carrés des parallélogrammes à angles donnés, ou semblables à un parallélogramme donné. La nature de cette transformation de l'algèbre ressort du mémoire cité; mais de même que plusieurs autres transformations particulières elle aurait mérité de faire, dans un chapitre particulier, l'objet d'une étude plus détaillée. Et c'est pourquoi j'y reviens ici. Le but du présent travail est 1° de signaler les opérations qu'on savait exécuter pratiquement avant la réforme en question, et 2° de rendre compte du travail qu'on a dû exécuter soit pour en faire l'algèbre scientifique et générale sous forme géométrique qu'on trouve dans les Éléments d'Euclide, soit pour la développer ultérieurement.

Nous espérons par ce travail continuer les efforts des historiens modernes de la mathématique, en particulier ceux de Paul Tannery, pour établir un rapport de liaison entre la géométrie classique des Grecs et l'algèbre de Diophante. En même temps, nous croyons ainsi réfuter la conception qui regarde l'algèbre arabe comme une véritable innovation, ce que fait encore M. P. Boutroux.[1] L'algèbre arabe avait commencé par faire usage des procédés que les Grecs possédaient déjà avant l'ère platonicienne: elle s'était développée ensuite en s'appropriant successivement les vues des grands auteurs scientifiques de la Grèce, ce qui ne veut pas dire pourtant que pendant leur travail d'habiles continuateurs les Arabes ne soient pas parvenus à exprimer

[1] Dans son intéressant ouvrage intitulé: Les Principes de l'Analyse mathématique, exposé historique et critique, I, Paris 1914.

les pensées déjà cultivées par les Grecs d'une manière qui à pu préparer les véritables progrès qui font ensuite l'honneur de la Renaissance.

II. L'arithmétique et l'algèbre avant Platon.

La revendication pour les Pythagoriciens d'une algèbre comprenant au moins la solution d'équations numériques du second degré appartient à Paul Tannery, qui a allégué de bonnes raisons en sa faveur dans son Mémoire intitulé: De la solution géométrique des problèmes du second degré avant Euclide.[1] Si je crois y pouvoir ajouter encore quelque chose, ce sera en m'occupant de la différence de vues et d'objectifs qui sépare la pratique des plus anciens mathématiciens et la conception scientifique d'Euclide.

Paul Tannery dit (p. 269 des Travaux scientifiques I): »Les quatre premiers livres (d'Euclide) représentent, comme fond et comme forme, la matière élaborée avant Eudoxe, les deux suivants supposent le travail de ce dernier, mais comme forme seulement, car il n'est pas douteux que la totalité des théorèmes qu'ils renferment n'aient dû être connus bien auparavant, sinon rigoureusement démontrés en dehors de l'hypothèse de la commensurabilité.« Les lecteurs de mon mémoire sur la Réforme sauront que je n'accepte pas une telle distinction entre les différentes parties d'Euclide, dont l'ordonnance est plutôt dictée par des considérations scientifiques que dues à des raisons historiques. Certainement je crois, moi aussi, que pratiquement les opérations algébriques ont été rattachées dès leur origine aux symboles géométriques; mais l'ancienneté de la nomenclature géométrique de l'arithmétique grecque, l'emploi arith-

[1] Mémoire de la Société des sciences physiques et naturelles de Bordeaux 1882, t. IV; Mémoires scientifiques I, p. 254.

métique par exemple des mots rectangle, carré etc., montre selon moi qu'il n'y faut pas voir l'application aux nombres d'une géométrie déjà faite, mais seulement un essai de créer un moyen de retenir les nombres et les opérations qu'on leur fait subir. Rectangles et carrés sont en effet, malgré leur nature complexe au point de vue logique, les notions intuitives les plus primitives par lesquelles on a commencé la géométrie, qui s'est plutôt développée pour satisfaire des besoins arithmétiques. C'est ainsi que la nature universelle des nombres a pu permettre de faire l'addition numérique d'un rectangle et d'un de ses côtés, ce que font sans scrupule les auteurs des travaux qui portent le nom de HERON et aussi DIOPHANTE, dans leurs exercises numériques, — et ce qu'on a fait probablement dès le commencement: on l'a fait en pensant seulement aux nombres que représentent ces figures, alors que le même emploi aurait constitué une hérésie dans la géométrie scientifique et dans ses applications aux démonstrations de propositions algébriques générales. De même, il est impossible qu'avant PLATON et EUCLIDE les manipulations géométriques dont on se servait pour exprimer la solution d'une équation du second degré, aient abouti à une construction méritant par sa dépendance de postulats expressément énoncés d'être regardée comme fournissant l'exacte valeur d'une racine générale; car les énoncés formels des postulats ne datent que d'un successeur de PLATON, probablement de MÉNECHME, et même l'usage pratique du compas à dessiner, d'HYPSICLÈS. Si, par exemple, EUCLIDE réduit dans II, 5, sous forme géométrique, l'équation

$$ax - x^2 = B, \qquad (1)$$

où a et x sont des droites limitées, B la surface d'une figure rectiligne, et où ax et x^2 sont représentés par un rectangle et un carré, à

$$\left(\frac{a}{2} - x\right)^2 = \left(\frac{a}{2}\right)^2 - B. \qquad (2)$$

nous ne doutons nullement que cela ne soit un emprunt à des connaissances antérieures; mais dans la géométrie scientifique d'Euclide cela fournit le moyen de construire exactement les racines de l'équation donnée. Jusqu'alors, l'équation (2) avait fourni le moyen de calculer les racines de l'équation (1), ou bien les nombres dont la somme sera a et le produit B si les valeurs numériques de a et B sont données. En effet, le calcul de $\left(\frac{a}{2}\right)^2 - B$ n'aura alors présenté aucune difficulté, et le calcul approché de ce que nous appelons la racine carrée d'un nombre, a été connu à cette époque depuis longtemps, quels qu'aient été d'ailleurs les moyens employés pour l'exécuter. L'idée de l'irrationalité de racines carrées n'était en effet qu'une réponse à la question de savoir si l'insuffisance des approximations résultait d'une nécessité absolue ou seulement de la défectuosité des méthodes jusqu'alors employées. Il est bien possible toutefois qu'on se soit efforcé de se servir pratiquement de constructions assez bonnes pour remplacer l'extraction approchée d'une racine carrée par quelque détermination d'un segment qu'on pouvait mesurer ensuite; mais on s'est servi alors d'autres instruments qu'Euclide, notamment d'un gnomon matériel au lieu du compas à dessiner (voir mon mémoire sur la Réforme p. 64). Du reste, le mesurage a pu se faire avec une assez grande précision au moyen de la même méthode qui sert à la détermination du plus grand facteur, méthode que connaissaient du moins Théodore et Théétète, et dont l'ancienneté est attestée par le fait que pour la désigner on s'est servi de bonne heure d'un nom particulier: ἀνταναίρεσις ou ἀνθυφαίρεσις (voir Réforme p. 108).

Les sources de nos connaissances de la mathématique préeuclidienne sont d'une part des rapports directs sur son état et même des extraits tels que celui du fragment d'Hippocrate de Chios, et d'autre part des conclusions indirectes que permet la manière dont Euclide et les auteurs suivants traitent les questions mentionnées dans les dits rapports. C'est au moyen de la première catégorie de sources que P. Tannery a établi que la connaissance de la solution d'équations numériques du second degré remonte aux Pythagoriciens. Avec son sens critique ordinaire, il distingue bien entre la valeur des différentes sources[1], et en particulier de celles où Proclus a puisé les renseignements qu'on lui doit. Il est vrai pourtant qu'on n'a pas à présent la même confiance qu'autrefois dans les comptes rendus de Proclus: même dans les cas où il puise aux meilleures sources, on le suspecte de les entremêler de traditions néopythagoriques[2]; mais, selon moi, on méconnaît peut-être la valeur de ces traditions. Bien que ces traditions se rattachent plutôt au mysticisme qu'à la science des anciens Pythagoriciens, l'emploi mystique de faits mathématiques témoigne d'une connaissance antérieure des mêmes faits, connaissance qui devait alors remonter au maître lui-même ou à ses premiers disciples. Ainsi mainte tradition représenterait à la fois l'un et l'autre. L'hésitation qu'on éprouve à attribuer à Pythagore la connaissance de cer-

[1]) Il y a pourtant un des arguments de P. Tannery auquel je ne puis adhérer. Il rappelle (p. 275 des Mém. sc.) que la construction d'une des lunules d'Hippocrate dépend de la solution géométrique d'une équation mixte du second degré. Or la construction en question est une intercalation (νεῦσις), et il n'y a pas de raison pour croire que déjà Hippocrate eût préféré à une exécution mécanique de cette opération une construction légitime d'après les principes d'Euclide.

[2]) Voir en particulier Eva Sachs: Die fünf platonischen Körper (Philologische Untersuchungen XXIV, 1917).

taines vérités ou procédés provient aussi, sans doute, souvent de ce fait qu'à présent une telle connaissance nous paraît inséparable de la démonstration euclidienne, que certainement on aurait tort de reporter à Pythagore; mais il ne faut nullement croire qu'à l'origine de la mathématique l'observation des faits et l'invention des procédés ait suivi le même cours que demande leur établissement formel sur les principes adoptés depuis Euclide. Les démonstrations bien formulées ont en général été précédées d'une découverte des faits et des procédés qui ne présupposait au fond que du bon sens intuitif.

Ce sont des connaissances d'ordre arithmétique et algébrique, acquises anciennement de cette manière, que je cherche à tirer des sources indirectes dont je viens de parler. J'ai montré, à la page 62 de mon Mémoire cité sur la Réforme, en quoi a consisté la généralisation de procédés déjà connus qui fait l'objet des démonstrations du livre II d'Euclide et quels ont été par conséquent les procédés analogues dont on disposait avant lui. J'y reviendrai tout à l'heure; mais d'abord je profiterai de l'Arithmétique de Nicomaque pour en tirer des formes encore plus anciennes. Un tel usage d'auteurs comme Nicomaque et Théon de Smyrne me paraît tout indiqué. En effet, dans un ouvrage où il se propose d'exposer les mathématiques nécessaires pour comprendre les endroits mathématiques dans les œuvres de Platon, le dernier de ces auteurs a dû rendre compte de doctrines connues avant Euclide. Et pas plus que lui, Nicomaque ne peut être regardé comme un mathématicien original. C'est particulièrement en parlant de nombres figurés qu'il rend compte de doctrines connues avant lui. On le voit immédiatement dans le chap. 16 de son livre II. En parlant des différentes dénominations de nombres à trois

dimensions, il dit que quelques auteurs en emploient une, d'autres une autre. L'emploi des nombres à trois dimensions n'était donc pas nouveau à son époque, et celui des nombres polygonaux plans a dû le précéder. Alors, comme pour cette doctrine il ne fait aucun usage d'Euclide, dont les vues abstraites seraient trop scientifiques pour les lecteurs de Nicomaque et de Théon, il faut croire que leur arithmétique géométrique, qui fait usage de points pour représenter les unités d'un nombre entier, et de l'ordonnance géométrique de ces points pour étudier les propriétés de ces nombres, avait été en usage avant l'algèbre géométrique d'Euclide, qui représente les quantités générales par des longueurs de droites et des surfaces de figures planes. L'idée de cette algèbre géométrique a donc été provoquée par les résultats qu'on avait su obtenir de l'arithmétique géométrique.

L'étendue du système de dénominations de différents nombres plans et solides que nous apprend Nicomaque, le rend même probable que les anciens mathématiciens, auxquels on doit originairement la notion de ces nombres, en ont fait des applications plus remarquables que celles qu'il nous communique, et leur usage peut même avoir préparé la véritable étude géométrique des mêmes figures[1]; mais ici nous devons nous borner aux applications les plus simples de nombres plans dont on peut constater plus directement un usage antérieur à l'époque de Platon et d'Euclide.

Un nombre triangulaire, dont les unités sont représentées par des points ordonnés sur des rangs parallèles contenant respectivement 1, 2, 3, . . ., n points, est leur somme $1 + 2$

[1] Dans une Note intitulée: L'Arithmétique géométrique des Grecs et des Indiens (Bibliotheca mathematica 5^3 (1904)) je me suis occupé des applications de cette nature.

$+3 \ldots + n$. Il est facile de voir par une juxtaposition de deux triangles de cette nature que la somme est égale à $\frac{1}{2} n(n+1)$. Nicomaque n'énonce pas toutefois directement ce résultat, mais il se réfère à des nombres carrés n^2 et à une espèce de rectangles qu'on appelait hétéromèques, et qu'on forme de n rangs dont chacun contient $n+1$ points; nous écrivons $n(n+1)$. La formation géométrique montre que les différences de deux nombres carrés consécutifs $\big((n+1)^2 - n^2 = 2n+1\big)$ sont les nombres impairs, et que les différences de deux nombres hétéromèques consécutifs $\big(n(n+1) - (n-1)n = 2n\big)$ sont les nombres pairs (Nicomaque II, 19). Les différences sont, en effet, les gnomons formés d'un rang horizontal et d'un rang vertical de la plus grande des figures. Il en résulte que le carré $(n+1)^2$ est la somme des nombres impairs $(1+3+5+\ldots+2n+1)$, le nombre hétéromèque $n(n+1)$, celle des nombres pairs $2+4+6+\ldots+2n$. Comme du reste le rectangle représentant le nombre hétéromèque $n(n+1)$ se compose de deux triangles au côté n, le dernier résultat contient la sommation $1+2+\ldots+n = \frac{1}{2} n(n+1)$.

Voilà un minimum de la connaissance et de l'application des nombres figurés qui datent déjà de l'époque des Pythagoriciens. En effet, le neveu de Platon, Speusippe s'en est occupé dans son livre sur les nombres pythagoriques, où il désigne aussi les progressions arithmétiques comme connues sous le nom de la première des progressions, ce qui fait aussi supposer la connaissance de progressions géométriques sous le nom de la seconde progression.[1]

Un point où les différents rapports se trouvent unanimes

[1] Voir P. Tannery: Un fragment de Speusippe (Annales de la Faculté des Lettres de Bordeaux (1883) t. V, ou Mémoires scientifiques (I, p. 281—289.

est l'ancienneté de la connaissance du fait que les carrés sont les sommes des nombres impairs; THÉON DE SMYRNE y revient trois fois. C'est de cette connaissance qu'on a déduit la solution de l'équation $x^2 + y^2 = z^2$ attribuée à PYTHAGORE; il suffit, en effet, d'égaler un gnomon, $(n+1)^2 - n^2$, qui est égal au nombre impair $2n+1$, à un carré impair quelconque r^2. Alors $n = \frac{1}{2}(r^2-1)$, et on obtient la solution $x = r$, $y = n = \frac{1}{2}(r^2-1)$, $z = n+1 = \frac{1}{2}(r^2+1)$. On aura de même la solution que, sans doute à tort, on attribue à PLATON, et qui doit être en tout cas préeuclidienne. On y fait usage d'un gnomon contenant deux rangs de points. Un tel gnomon sera la différence de deux carrés dont nous pouvons dire que les côtés contiennent respectivement $n+1$ et $n-1$ points. Le gnomon contiendra alors $4n$ points, ce que montre immédiatement une figure, tandis que l'algèbre actuelle le constate à l'aide du calcul: $(n+1)^2 - (n-1)^2 = 4n$. En remplaçant ici n par un carré quelconque r^2 on obtient la solution $x = 2r$, $y = r^2-1$, $z = r^2+1$.

Aux carrés et gnomons formés de points, qui ont suffi ici, EUCLIDE a substitué dans II, 4 et 7 de véritables figures géométriques. Si, pour traduire ces deux théorèmes, qui se rattachent à la même figure, en une formule algébrique moderne, on appelle le côté de l'un des carrés a et la largeur du gnomon b, on voit que les deux théorèmes expriment la même chose que les formules

$$(a \pm b)^2 = a^2 \pm 2ab + b^2. \qquad (1)$$

A l'endroit cité, EUCLIDE, qui vient de démontrer le théorème de PYTHAGORE, introduit et démontre ces deux théorèmes pour en faire usage dans la démonstration de son extension à un triangle quelconque que contiennent les théorèmes (II, 12 et 13) relatifs aux côtés et à la projection

d'un côté sur un autre (le théorème du cosinus); l'usage correspond à celui que dans la même démonstration on fait à présent des deux formules (1).

Les propriétés de la figure de gnomon se traduisent aussi, si les côtés sont appellés a et b, par la formule algébrique

$$a^2 - b^2 = (a+b)(a-b),$$

traduction qui correspond à l'usage qu'Euclide fait de la même figure dans son lemme à X, 28 pour obtenir la solution la plus générale de l'équation pythagorique. En substituant aux côtés d'un rectangle égal au gnomon (que nous venons d'appeler $a+b$ et $a-b$) deux nombres semblables (nombres proportionnels à deux carrés) m^2c et n^2c, on trouve ainsi la solution générale

$$x = mnc, \quad y = \tfrac{1}{2}(m^2 - n^2)c, \quad z = \tfrac{1}{2}(m^2 + n^2)c. \quad (3)$$

Cette solution comprend les deux dont nous avons déjà parlé: pour $c = 1$, $m =$ un nombre impair quelconque et $n = 1$, celle qu'on attribue à Pythagore, pour $c = 2$, $m =$ un nombre quelconque, $n = 1$, celle qu'on attribue à Platon. Elles résulteraient de II, 4 en y donnant au gnomon respectivement les largeurs 1 et 2. Néanmoins Euclide démontre dans II, 8 un théorème particulier, dont il ne fait aucun usage lui-même, et dont le seul but a dû être de donner une forme géométrique et plus exacte selon lui à la démonstration dont on s'était servi jusqu'alors pour établir la solution de Platon. Dans la même traduction algébrique que dans notre formule (1) nous avons appliquée aux théorèmes II, 4 et 7, le théorème énonce que

$$(a+b)^2 - (a-b)^2 = 4ab.$$

En y remplaçant a par un nombre carré, b par 1, on aura la même démonstration que nous venons d'attribuer à l'arithmétique préeuclidienne.

II, 8 n'est pas la seule des dix premières propositions du livre II d'Euclide dont l'unique but visible est de conformer une démonstration connue antérieurement aux principes suivis dans ses Eléments. En effet, Théon de Smyrne nous fait connaître une série de valeurs approchées de $\sqrt{2}$, à savoir $\frac{3}{2}$, $\frac{7}{5}$, $\frac{17}{12}$, .., qui non seulement étaient connues par Platon, mais qui remontent jusqu'aux Pythagoriciens[1]: c'est de leur approximation à $\sqrt{2}$, qu'Euclide donne des démonstrations exactes dans II, 9 et 10. Il ne me paraît pas possible de tirer de ces démonstrations les procédés par lesquels on avait avant lui trouvé la série en question et vérifié son approximation à $\sqrt{2}$. Toutefois les valeurs approchées sont les mêmes qu'on obtient en développant $\sqrt{2}$ en fraction continue. Or nous avons parlé de l'ancienneté de la détermination de la plus grande mesure commune de deux nombres, procédé qu'on peut appliquer aussi à deux quantités représentées par des segments de droites, ce qu'on fera en exécutant géométriquement les soustractions, et on sait que le même procédé conduit au développement en fraction continue. Il est donc probable que c'est par l'application des mêmes procédés au rapport de $\sqrt{2}$ à 1, représenté par celui de la diagonale d'un carré à son côté, qu'on est parvenu à la même suite de valeurs approchées successives. En faveur de cette hypothèse on peut citer le fait que selon Aristote (158b 29) l'identité des »antanairèses« (voir p. 12) de deux rapports » servi plus tard de critère de leur égalité.[2] C'est sans doute d'abord par induc-

[1] Voir le commentaire de Proclus à la République de Platon, ed. Kroll 1901.

[2] Quant à la forme d'une telle application des procédés, on en est réduit toutefois à des hypothèses incertaines. En tout cas, comme dans le développement ordinaire de $\sqrt{2}$ en fraction continue, on a dû invoquer le fait que $(\sqrt{2}-1)(\sqrt{2}+1) = 1$, ce que montre la figure d'un

tion qu'on a trouvé la loi générale de la formation du convergent $\frac{y_1}{z_1}$ qui suit au convergent $\frac{y}{z}$, $y_1 = y+2z$, $z_1 = y+z$. C'est l'utilité de tels convergents qu'Euclide démontre dans les propositions II, 9 et 10, qui expriment, sous sa forme géométrique habituelle, que

$$(y+2z)^2 - 2(y+z)^2 = -(y^2-2z^2).$$

Alors, si $y^2-2z^2 = +1$, on aura $y_1^2-2z_1^2 = -1$, et, à cause de la croissance du numérateur et du dénominateur, l'approximation de $\sqrt{2}$ deviendra de plus en plus fine.

Le but visé par Euclide dans ses propositions II, 8, 9, 10 est donc de montrer combien les principes qui font la base du système géométrique réalisé dans ses Éléments se prêtent bien aussi à la démonstration exacte de vérités arithmétiques qu'il regarde comme trop connues pour qu'il ait besoin de rendre compte de leurs applications numériques. II, 4 et 7 sont introduits à cause de l'usage qu'Euclide va en faire ensuite dans son système géométrique; mais nous avons vu que leur usage arithmétique était bien connu avant lui. Telle doit être la nature de tous les dix premiers numéros du livre II d'Euclide, qui déjà par leur forme font un groupe particulier où l'on ne trouve pas la déduction successive des précédents. 1—3 servent seulement d'introduction; mais 5 et 6 contiennent les démonstrations euclidiennes de transformations algébriques qui ont dû être connues avant Euclide. Nous avons déjà dit que 5 contient, sous forme géométrique, celle que nous exprimons à présent par

gnomon. Si on substitue un convergent $\frac{y}{z}$ à la dernière $\sqrt{2}$, on aura pour la première le convergent suivant $\frac{y_1}{z_1} = \frac{y+2z}{y+z}$. C'est probablement par le même procédé que Théodore a pu démontrer l'irrationalité de $\sqrt{3}$, $\sqrt{5}$.., $\sqrt{18}$. Voir mon mémoire: Sur les livres arithmétiques d'Euclide (Bulletin de l'Académie de Danemark 1910).

$$ax - x^2 = \left(\frac{a}{2}\right)^2 - \left(\frac{a}{2} - x\right)^2.$$

De même 6 contient celle que nous exprimons à présent par

$$ax + x^2 = \left(\frac{a}{2} + x\right)^2 - \left(\frac{a}{2}\right)^2.$$

Voilà les artifices qui servent à la solution d'équations du second degré, connue déjà par les Pythagoriciens, qui, eux, ont dû la trouver moyennant les mêmes figures dont se sert Euclide, mais sans ses constructions bien formulées de ces figures et des solutions. Tant qu'on se bornait à des cas où a et x sont des nombres entiers, a un nombre pair, on a pu former ces figures avec des points représentant les unités; mais, en rencontrant des équations à racines irrationnelles, on a dû éprouver la nécessité de substituer aux points des figures géométriques dont les côtés représentent des quantités simples, les surfaces, des produits.

Grâce à cet artifice, on a pu résoudre des équations numériques qui se sont présentées, par exemple, à la recherche de deux nombres dont le produit et la somme, ou la différence, sont donnés. On a pu faire cela pratiquement avant de bien formuler les règles pour la formation et la solution des équations. D'ailleurs Euclide ne donne pas encore ici de tels énoncés: il les réserve pour son livre VI, où la théorie générale des proportions lui permet de leur donner une plus grande généralité; mais cela ne l'empêche pas d'en faire déjà une application pratique pour obtenir une solution dont il a besoin avant le dit livre. C'est ainsi que dans la proposition II, 11 il applique II, 6 à la solution de l'équation

$$a(a - x) = x^2,$$

qui donne

$$\left(x + \frac{a}{2}\right)^2 = \left(\frac{a}{2}\right)^2 + a^2.$$

Si a est un nombre donné, on en pourra déduire, au moyen d'une valeur approchée de $\sqrt{5}$, une valeur approchée de x; mais Euclide s'en sert pour construire exactement x, a étant alors un segment donné, et il en fait usage dans le livre IV pour construire un pentagone régulier. Peut-être les Pythagoriciens ont-ils su de même construire la racine au moyen du théorème de Pythagore, et ils en ont peut-être fait le même usage pratique sans attacher encore aux constructions géométriques la même importance théorique que leur attribue Euclide.

Euclide nous apprend donc dans son livre II l'une des manières dont les Grecs avaient représenté déjà avant lui les opérations algébriques, en même temps qu'il donne à ces représentations la forme que demandent ses nouveaux principes rigoureux, et dont il va se servir dans la suite de ses Éléments. L'autre organe des opérations algébriques des Grecs est l'usage des proportions, dont Euclide s'occupe de deux manières différentes correspondant à deux définitions différentes. Dans le livre V, il prend les principes d'Eudoxe pour point de départ d'une théorie générale également applicable à des quantités commensurables et incommensurables, théorie qui fait ensuite, dans le livre VI, à côté de l'usage des angles, la base solide de la théorie des figures semblables. Dans le livre VII, un subtil usage d'une définition arithmétique — et par là moins générale — des proportions, fournit à Euclide, qui suit à cet égard probablement des principes inventés par Théétète, une base solide des démonstrations des théorèmes sur la divisibilité des nombres composés, théorèmes nécessaires pour établir rigoureusement, dans le livre X, les critères de la commensurabilité des radicaux.[1] Le fait qu'on avait

[1] Voir mon mémoire sur les livres arithmétiques d'Euclide (Bulletin de l'Académie de Danemark, 1910).

recours aux proportions pour réaliser deux buts scientifiques différents, et d'autre part le nombre des propriétés des proportions qu'on avait à sa disposition, et qu'à ces effets on déduisait rigoureusement de définitions bien formulées, prouve que du moins l'usage pratique des proportions a été familier aux Grecs avant l'époque des dits savants. Probablement on avait originairement démontré une partie de ces propriétés pour les cas où les termes des proportions étaient des nombres entiers, et on les aura généralisés ensuite par induction; mais on a eu un moyen encore plus intuitif pour trouver ces propriétés. En effet, psychologiquement, la conception de la similitude est plus primordiale à l'homme que la notion des proportions, tandis que depuis Euclide la théorie des proportions fait inversement la base logique de l'étude des figures semblables. Avant Euclide, il a été assez naturel de tirer plusieurs des propriétés des proportions de la seule considération de figures visiblement semblables. (Voir la Section IX de mon mémoire sur la Réforme.) En considérant, par exemple, un triangle avec une transversale parallèle à un de ses côtés, on aura été conduit immédiatement à admettre les transformations suivantes d'une proportion (Euclide V, 16—19):

de $$a : b = c : d$$

en $$a : c = b : d$$

ou $$(a \pm b) : b = (c \pm d) : d$$

ou $$a : b = a \pm c : b \pm d.$$

Aux démonstrations de ce genre Euclide a substitué de plus rigoureuses dans son livre V et dans les livres VII—IX. Ainsi il s'est créé des fondements essentiels pour ses propres recherches géométriques; mais en même temps, dans les trois livres arithmétiques comme dans le livre II, Euclide

profite de l'occasion pour étendre la même rigueur aux démonstrations d'autres propositions dont on avait fait avant lui un usage arithmétique.

C'est ce qui explique qu'après avoir donné la théorie générale de la divisibilité des nombres composés, dont il avait un besoin immédiat dans son système, il s'occupe, vers la fin du livre IX, particulièrement des nombres pairs et impairs, impairement pairs, etc. et énonce à ce sujet des théorèmes qui paraissent assez triviaux auprès de ceux qui précèdent. La division des nombres en pairs et impairs, à laquelle se joignait de nouvelles catégories suivant la divisibilité par quatre, etc., avait, en effet, depuis Pythagore une grande popularité, et nous apprenons de l'ouvrage de Nicomaque que cette popularité s'est conservée indépendemment des vues plus générales qu'Euclide avait prises avant de s'occuper de cette matière assez spéciale. Il est assez probable que la belle démonstration (IX, 20) de l'infinité de la série des nombres premiers a été également connue avant Euclide. Dans cette hypothèse, ce théorème, dont l'intérêt est purement arithmétique, doit sa place, dans ses Éléments, à la circonstance qu'on y fait usage d'un théorème sur la divisibilisé des nombres qu'Euclide avait dû démontrer rigoureusement (VII, 31).

Une grande place est accordée dans ces livres à la théorie des puissances sous forme de théorie des proportions continues. S'il y a $n+1$ termes dans une telle proportion

$$a : b = b : c = c : d = \ldots = e : f$$

le rapport $\frac{f}{a}$ du dernier terme f au premier terme a sera ce que nous appelons aujourd'hui la $n^{\text{ième}}$ puissance du rapport $\frac{b}{a}$ du deuxième terme b au premier a. Si le premier terme est 1, le second un nombre entier, on aura ainsi les puissances du nombre entier. Inversement, l'extrac-

tion de la $n^{\text{ième}}$ racine de $\frac{f}{a}$ est représentée comme l'insertion de $n-1$ termes d'une proposition continue commençant par a et finissant par f. Cette représentation des puissances et des radicaux remonte au moins à Hippocrate de Chios, qui a réduit la duplication du cube à l'insertion de deux moyennes proportionnelles. Il a donc reconnu l'identité des deux manières dont les anciens représentaient une $3^{\text{ième}}$ puissance: 1° géométriquement, comme un cube, 2° au moyen d'une proportion continue à 4 termes. Euclide, de son côté, donne dans ses livres arithmétiques un développement assez étendu de cette théorie des puissances. Néanmoins, dans le livre X, il ne se soucie pas d'en tirer pour les radicaux d'autres fruits que ceux qui sont relatifs aux racines carrées, seules racines dont il ait besoin dans ses recherches géométriques propres, où il se borne aux opérations qui ne demandent que l'usage de la droite et du cercle ou d'équations du second degré; mais la généralité de sa théorie des puissances, et en particulier de ses théorèmes VII, 27 et VIII, 2 lui aurait permis d'étendre aux radicaux de tous les ordres les critères (X, 9) de la rationalité de racines carrées. Aussi voit-on, du dialogue de Platon qui porte le nom de Théétète, que ce savant s'est occupé de la rationalité de racines cubiques. Il est donc probable qu'une partie, au moins, des résultats relatifs aux proportions continues lui est due; mais sans doute Euclide en a élaboré ultérieurement la théorie.

La fin du livre IX contient une application de la théorie des puissances à la détermination d'un nombre dit parfait, qui semble remonter en partie aux Pythagoriéiens, tandis qu'on y reconnaît aussi une généralisation de la démonstration qui porte plutôt l'empreinte d'Euclide. Est »parfait« selon sa définition VII, 22 tout nombre qui égale la somme

de ses parties aliquotes, et le théorème IX, 36 énonce que, dans le cas où la somme $s = 1+2+2^2+\ldots+2^n$ est un nombre premier, le nombre $s \cdot 2^n$ est parfait. En effet, ses parties aliquotes seront alors

$$1,\ 2,\ 2^2,\ \ldots,\ 2^n,\ s,\ 2s,\ \ldots,\ 2^{n-1}s,$$

et leur somme

$$s+s(1+2+2^2+\ldots+2^{n-1}) = s+s(2^n-1) = s\cdot 2^n.$$

Tout cela peut très bien avoir été connu par les Pythagoriciens. La notion de nombre parfait se rattache à des propriétés qui leur intéressaient, et le nombre 6 en offrait un simple exemple. Nous avons déjà parlé de leur connaissance de séries géométriques. Les puissances du nombre 2 en offraient l'exemple le plus simple, et Philolaus attribue aux planètes des distances du feu central qui sont exprimées par les puissances de 3. Vu que $2^r-2^{r-1} = 2^{r-1}$, ils ont très bien pu trouver par induction, et voir confirmée ensuite, la règle concernant la sommation des puissances de 2. Alors, en remarquant que le nombre parfait le premier connu 6 peut s'écrire $(1+2)2$, et en essayant d'en trouver d'autres, on a pu être conduit à observer que $28 = (1+2+2^2)2^2$ en constitue un également. En continuant, on aura vu que $(1+2+2^2+2^3)2^3$ n'en est pas un parce que 15 n'est pas premier, mais que $31 \cdot 2^4$ au contraire en est un, et ainsi on peut avoir été amené à établir la règle générale, en partie par les mêmes considérations auquelles nous avons pu donner ici des expressions plus simples qu'il ne leur était possible; mais le défaut de ces moyens modernes a été récompensé par la considération directe des nombres eux-mêmes qu'on a retenue avec plus d'assiduité qu'on n'a besoin de le faire aujourd'hui.

S'il en est ainsi, Euclide a fait ici la même chose que dans son livre II, en faisant dépendre l'établissement d'un

résultat arithmétique qui avait déjà une certaine popularité d'un théorème qui découle des vues générales qu'il a dû prendre pour réaliser ses propres buts. Ce théorème (IX. 35), qui est à nos yeux beaucoup plus important que ne l'est la formation de nombres parfaits, exprime la sommation générale d'une série géométrique. Il le demontre de la manière suivante. En appliquant des transformations connues aux proportions

$$\frac{a}{b} = \frac{b}{c} = \frac{c}{d}$$

on aura $$\frac{b-a}{a} = \frac{c-b}{b} = \frac{d-c}{c} = \frac{d-a}{a+b+c},$$

et sans qu'Euclide ait eu besoin de le dire expressément, on voit que le même raisonnement peut être appliqué quel que soit le nombre des termes de la série. Dans le cas où $a = 1$, $b = 2$, cette démonstration serait à peu près la même que nous venons d'attribuer aux Pythagoriciens, qui n'ont pas, pour l'effectuer, eu besoin de proportions.

Rappelons encore un exemple qui montre qu'avant d'avoir à sa disposition des moyens techniques propres aux questions dont on s'occupait, on a su se servir de la langue courante pour les exprimer et pour en donner les solutions. Il nous est rapporté que le Pythagoricien Thymaridas[1] a exposé les équations que nous écririons

$$x_1 + x_2 + \ldots + x_n = s, \quad x_1 + x_2 = a_1, \quad x_1 + x_3 = a_2, \ldots,$$
$$x_1 + x_n = a_{n-1}$$

et leur solution

$$x_1 = \frac{a_1 + a_2 + \ldots + a_{n-1} - s}{n-2}$$

[1] Dans un mémoire inséré aux Annales de la Faculté des Lettres de Bordeaux (1881), t. III p. 101—104 (Mémoires scientifiques t. I p. 106—110) P. Tannery l'a rendu probable que Thymaridas appartenait aux *anciens* Pythagoriciens.

de la manière suivante: Si des quantités (données et inconnues font une quantité donnée, et que l'une d'entre elles ajoutée à chacune des autres fasse une somme donnée, alors la somme des couples moins la somme donnée appartiendra, totalement dans le cas de trois nombres, de moitié dans le cas de quatre, du tiers dans le cas de cinq, du quart dans le cas de six, etc. à celle qui est ajoutée à chacune des autres.

Nous avons montré qu'avant l'ère platonicienne l'algèbre grecque avait à sa disposition les deux représentations techniques suivantes: 1° les surfaces de rectangles et carrés et 2° les proportions simples et continues, et qu'en manipulant pratiquement les premières on savait résoudre les équations du second degré, et que grâce aux dernières on savait représenter les puissances et les racines. Dans ce qui suit, nous nous occuperons 1° de la généralisation de ces opérations et de leur réduction à des vues plus exactes commencées par les contemporains de Platon, et 2° de leurs applications plus générales continuées aussi après Euclide. Étant donné leur origine, il est évident que la forme plus générale et géométrique qu'on leur donna à cette époque n'a pas fait oublier aux auteurs de ces progrès les applications simples et numériques qu'on peut en faire. Entre leurs mains, ces moyens n'ont donc pas servi seulement à étendre le domaine de la géométrie pure, mais ils sont restés aussi pour eux des instruments servant d'une façon générale à résoudre les questions que nous résolvons à présent au moyen de l'algèbre littérale. C'est pourquoi je regarde ici leur développement comme un développement de l'algèbre, qui se sert de nos jours d'autres symboles.

D'un autre côté, après les Arabes, qui sont à ces égards les élèves immédiats des Grecs, l'adhésion à l'algèbre grecque

est rendue plus pénible par les vues abstraites que prend toujours Euclide; c'est l'abstraction de ces vues qui a fait naître chez plusieurs historiens modernes des doutes sur la connexion de l'algèbre géométrique des Grecs avec la connaissance des emplois simples et numériques qu'on en peut faire, connaissance qui pourtant a dû précéder la grande abstraction dont ils marquent le point de départ.

III. Analyse algébrique sous forme géométrique.

Dans mon mémoire sur la Réforme j'ai montré, d'un côté, en quoi a consisté la généralisation et l'abstraction auxquelles on a, pendant l'intervalle de Platon à Euclide, soumis les procédés géométrico-algébriques dont avant cette époque on avait su faire des applications pratiques à des questions particulières. On substituait aux figures servant à marquer pratiquement les opérations et leurs résultats successifs des figures construites conformément à des postulats formellement énoncés, et on donnait à ces figures une généralité géométrique que, du reste, assez souvent les applications algébriques ne demandaient pas. Une généralisation plus importante fut obtenue au moyen d'une théorie générale des proportions, qui conférait aux compositions algébriques qui ne se faisaient pas par la simple formation de rectangles et de carrés cette propriété de pouvoir être appliquées indifféremment aux quantités rationnelles et irrationnelles dont jouissaient immédiatement les résultats de constructions géométriques.

D'un autre côté, je me suis occupé dans le chap. IV du mémoire cité, d'un procédé grâce auquel on a pu réaliser la réforme de la géométrie, et en même temps celle de l'algèbre qui se servait de symboles géométriques: je veux parler de l'*analyse* des notions et des vérités qu'on possédait

déjà. Dues en grande partie à l'intuition, celles-ci ne jouissaient pas de la simplicité qu'on doit demander aux points de départ d'un développement logique. Il fallait y appliquer une analyse pour dissoudre les notions et vérités connues en leurs »éléments« jusqu'aux premiers dont on pouvait composer ensuite synthétiquement un système contenant des démonstrations exactes de toutes les vérités géométriques, y compris celles dont on avait déjà eu une connaissance préalable.

Cette analyse ainsi que la synthèse qui la suit sont au fond de la même nature que celles dont on a continué de faire usage ensuite pour réduire des théorèmes dont on n'a encore aucune démonstration formelle à des vérités déjà démontrées (analyse théorique), et des problèmes qu'on n'a pas encore résolus, à des constructions déjà connues (analyse problématique), pour obtenir ensuite synthétiquement, par voie inverse, les démonstrations des théorèmes et les solutions des problèmes proposés. Cette »méthode analytique« qui sert aux progrès ultérieurs de la géométrie déjà bien constituée s'est développée pendant le même espace de temps où s'accomplissait la réforme grâce à laquelle la mathématique déjà existante est devenue science raisonnée. Cette méthode analytique donnait en particulier à l'usage des figures dont on s'était servi jusqu'alors pour représenter le petit nombre d'opérations *algébriques* qui étaient déjà connues, une étendue qui en faisait un utile organe pour de nouvelles recherches algébriques.

Aussi les premières applications de la méthode analytique qui nous soient rapportées ont-elles pour objet des recherches algébriques. En effet, dans le fragment historique inséré par Proclus dans la seconde partie de son Prologue (éd. Friedlein 67, 5—8) on lit qu'Eudoxe de Cnide

ταῖς τρισὶν ἀναλογίαις ἄλλας τρεῖς προσέθηκεν καὶ τὰ περὶ τὴν τομὴν ἀρχὴν λαβόντα παρὰ Πλάτωνος εἰς πλῆθος προήγαγεν καὶ ταῖς ἀναλύσεσιν ἐπ' αὐτῶν χρησάμενος.

ajouta trois nouvelles analogies aux trois anciennes, et fit progresser les questions relatives à la section soulevées par Platon et pour lesquelles il (Eudoxe) fit usage des *analyses*.

Depuis Bretschneider, on croit ordinairement que la section (ἡ τομή) dont il est question ici est la section d'une ligne en moyenne et extrême raison (*sectio aurea*). Tannery a émis quelques doutes à cet égard[1], auxquelles j'ai adhéré à la p. 36 de mon Mémoire cité. Toutefois, des arguments allégués par M^elle^ E. Sachs[2] m'ont fait changer d'avis, ce que je reconnais dans une note adjointe à mon mémoire (p. 170). M^elle^ Sachs fait aussi valoir plusieurs arguments en faveur de l'hypothèse ultérieure de Bretschneider suivant laquelle ce serait du travail d'Eudoxe sur ce sujet qu'on avait extrait la collection d'analyses et de synthèses relatives aux 5 premières propositions du livre XIII d'Euclide qui remplacent dans plusieurs éditions le texte ordinaire[3]. J'aurai l'occasion d'ajouter aux siens plusieurs arguments *pour* et *contre*, que je tire du contenu et de la forme géométrique de ces analyses et synthèses; mais il y a une conclusion que M^elle^ Sachs en tire et à laquelle je ne puis adhérer. Elle croit y voir la première détermination géométrique de cette section, nécessaire pour la construction du pentagone régulier.

En faveur de cette hypothèse elle fait remarquer la

[1] La Géométrie grecque p. 76.

[2] Dans son mémoire cité p. 97. Mes remarques suivantes se réfèrent à la p. 95, note 1.

[3] Heiberg les a renvoyées à son appendice I. (Éléments t. IV p. 364—377.)

complexité de cette solution dont on aurait pu se passer, selon elle, si l'on connaissait déjà la solution simple du même problème qu'on trouve dans Euclide II, 11. A ce sujet, il faut remarquer d'abord que la construction qu'on tirerait d'Euclide XIII, 1 ou 2 ne différerait pas de celle qu'il a donnée déjà dans le livre II, 11: ce dont on a besoin dans le livre XIII n'est pas la construction, mais l'expression algébrique de la plus grande partie (x) d'une droite (a) résultant de sa section en moyenne et extrême raison, celle qu'à présent nous écrivons $x = \frac{a}{2}\sqrt{5} - \frac{a}{2}$. Euclide la donne en énonçant sous la forme géométrique dont il se sert communément:

$$\left(x + \frac{a}{2}\right)^2 = 5\left(\frac{a}{2}\right)^2. \qquad (1)$$

Cette relation est une simple conséquence de II, 11 (voir p. 21). Aussi Eudoxe, si c'est bien lui, l'auteur des analyses et synthèses en question, ne se soucie-t-il pas de la déduire, mais seulement de la démontrer au moyen de la méthode analytique. Si la relation avait été entièrement inconnue jusqu'alors, une déduction aurait demandé une analyse problématique; mais l'auteur en a trouvé une démonstration au moyen d'une analyse théorique, et, devant commencer par regarder comme concédé ce qu'il s'agit de démontrer, une telle analyse a demandé une connaissance antérieure du résultat (ou du moins une supposition dont on voulait éprouver la vérité au moyen de l'analyse). Aussi l'analyse conservée est-elle une démonstration que la relation (1), regardée comme concédée, amène la relation

$$x^2 = a(a - x), \qquad (2)$$

qui exprime la section de a en moyenne et extrême raison. C'est seulement dans la synthèse suivante qu'on démontre qu'inversement cette dernière relation amène (1).

Tout cela s'accorde du reste très bien avec les expressions employées par Proclus, qui ne dit nullement qu'Eudoxe a imaginé la construction de la section, mais qu'il fit progresser les questions y relatives. Ce sont des questions semblables qui font l'objet des quatre autres propositions du commencement du livre XIII.

Le théorème XIII, 2 est l'inverse de XIII, 1. Sa démonstration euclidienne et la synthèse conservée sont donc essentiellement identiques à l'analyse qui avait servi à trouver la démonstration de XIII, 1. De même, la démonstration et la synthèse de XIII, 1 sont identiques à l'analyse appliquée à XIII, 2. Dans le théorème XIII, 3 on établit que si nous désignons par x_1 la plus petite partie de a, ou $a-x$:

$$\left(x_1+\frac{x}{2}\right)^2 = 5\left(\frac{x}{2}\right)^2,$$

théorème dont on trouve plus tard une application directe dans XIII, 16, mais d'où résulte aussi, selon 1, la continuation des sections en moyenne et extrême raison, donnant des segments de plus en plus petits. Dans XIII, 4 on démontre que l'équation (2) amène l'équation

$$a^2+(a-x)^2 = 3x^2. \tag{4}$$

Dans XIII, 5 on démontre la continuation des sections en moyenne et extrême raison donnant des segments de plus en plus grands: $a+x$ se décompose par la section en a et x. On montre, en effet, que l'équation (2) amène

$$x(a+x) = a^2. \tag{5}$$

Les démonstrations euclidiennes de 3, 4 et 5 sont essentiellement identiques aux synthèses qui y correspondent dans le fragment attribué à Eudoxe.

Ces théorèmes, de nature algébrique, sont, de même

que leurs démonstrations et les analyses et synthèses qui y correspondent, représentés au moyen des symboles géométriques ordinaires: rectangles et carrés, et ce que nous avons exprimé ici par des équations est énoncé comme égalités de surfaces; comme dans le livre II, on se sert de gnomons où il en est besoin pour déduire une égalité d'une autre. Pour l'invention de figures utiles à cet effet, on a été guidé par la connaissance antérieure de l'artifice utile pour résoudre une équation du second degré: c'est ainsi que, dans 1 et 2, on a été conduit à introduire des figures au côté $\frac{a}{2}$. De même qu'on commence de nos jours une recherche algébrique par la formation d'équations, où on peut regarder, à volonté, les différentes quantités représentées par des lettres comme données ou cherchées, on commençait alors par la formation des figures et en faisait un usage parallèle à l'usage actuel d'autres symboles. On avait ainsi une véritable *analyse algébrique sous forme géométrique.*

La collection conservée d'analyses et synthèses relatives aux 5 propositions d'Euclide XIII, doit avoir été déstinée à montrer l'emploi de cette analyse algébrique. Cela ressort déjà de l'introduction, où il est dit (Édition de Heiberg IV, p. 364): L'analyse est la supposition de ce qu'on cherche comme concédé, supposition d'où l'on parvient par ses conséquences à quelque chose dont la vérité est concédée. La synthèse est la supposition de ce qui est concédé; suppol'sition d'où l'on parvient par ses conséquences à [la vérité qu'on cherche].[1]

[1] Il s'est sans doute glissé une erreur dans le texte conservé qui se termine par les mêmes mots que l'explication de l'analyse: επί τι ἀληθὲς ὁμολογούμενον. Voir l'édition de Heiberg vol. IV p. 366, 2 et sa note relative à une correction faite par un éditeur antérieur.

Nous avons dit que toutes les analyses ont la forme dite *théorique*, tandis qu'il aurait paru naturel de commencer dans 1 par une analyse *problématique*, conduisant à la solution du problème de la section; les autres propositions sont de véritables théorèmes sur les segments qui résultent de la section. Abstraction faite de la forme, la démonstration donnée par Euclide de son théorème 1, comme aussi la synthèse identique à elle, contient du reste essentiellement ce que devait contenir une telle analyse problématique. Au cas où les analyses et synthèses, comme le supposent Heiberg et Tannery, seraient dues à un scholiaste postérieur, la forme théorique s'expliquerait simplement par le fait qu'Euclide expose toutes les cinq propositions sous forme de théorèmes. Ayant résolu, dans son livre II, une fois pour toutes le problème en question, il n'a pas bésoin d'y revenir; mais son intention évidente est d'avoir à sa disposition, pendant sa détermination suivante des côtés des différents polyèdres réguliers, les propriétés algébriques des segments résultant par la section en moyenne et extrême raison. Les déterminations des côtés poursuivent, en effet, elles aussi des buts algébriques malgré leur forme géométrique, ce qu'on voit par la discussion approfondie de la nature de leurs différentes irrationalités.

On voit donc que le groupe des 5 propositions, de même que le groupe des 10 premières propositions du livre II, est une collection de lemmes algébriques qu'Euclide insère pour en faire usage dans ce qui suit, et, de même que dans le livre II, il ne se borne pas à ceux dont il peut prévoir un emploi immédiat; mais il saisit l'occasion d'en démontrer plusieurs. Dans le livre II, cela s'explique par son désir de donner aussi aux méthodes usitées déjà pratiquement dans l'arithmétique la même rigueur géométrique

qu'à ses propres applications de procédés semblables. Dans les emprunts à l'algébre qu'il fait au commencement du livre XIII, il renonce à faire usage des résultats acquis dans ses livres précédents. Cette affirmation ne se rapporte pas seulement à II, 11: au lieu de renvoyer, comme il le fait dans son livre X, toutes les fois qu'il a besoin de la solution d'une équation du second degré, à la solution une fois acquise sous forme de paraboles de surfaces, Euclide construit dans son livre XIII, à chaque fois, le gnomon utile pour effectuer la solution. Les continuations des sections résultant des n^{os} 3 et 5 sont du reste des conséquences très simples de la théorie des proportions, et elles ont dû se présenter immédiatement au moment ou l'on a fait usage des »antanairèses« (voir p. 12) pour éprouver la commensurabilité d'un segment aux parties résultant d'une section en moyenne et extrême raison, ce qu'on n'a pu négliger, car ce moyen est celui que dans son livre X Euclide prescrit en premier lieu. C'est donc plutôt l'application qu'on y fait des figures servant à l'analyse algébrique que la nouveauté des résultats qui a porté Euclide à insérer cet emprunt à l'algèbre que contiennent les cinq propositions dont nous parlons et à en conserver la forme singulière. Cette forme s'explique dans l'hypothèse d'un emprunt fait à un auteur qui aurait eu pour but principal le développement de l'analyse algébrique sous forme géométrique; si cet auteur possédait une autorité comme celle d'Eudoxe, on comprendrait qu'Euclide ait conservé dans ses démonstrations à très peu près la forme des synthèses d'Eudoxe, ou même, comme on l'a cru, qu'il ait gardé telles quelles toutes les analyses et synthèses.

Même en attribuant les analyses et synthèses à Eudoxe et y voyant, avant tout, des modèles servant à illustrer la

méthode analytique, on ne doit pas s'étonner de n'y trouver que des analyses et synthèses »théoriques«. En effet, comme je l'ai montré dans le mémoire sur la Réforme (p. 39), ce fut l'élève d'Eudoxe Ménechme qui introduisit le premier l'usage systématique des problèmes; il n'y avait donc avant lui aucune raison pour distinguer entre analyses théoriques et problématiques. On remarquera aussi que cette distinction n'est pas faite dans les définitions, citées à la p. 34, qui servent d'introduction aux analyses et synthèses en question ici. Au temps d'Eudoxe, l'analyse avait plutôt encore la forme sous laquelle elle se prêtait, appliquée aux vérités déjà connues, à la transformation de l'ensemble de ces vérités en science raisonnée. Nous avons vu, en effet, que l'analyse conservée du théorème XIII, 1 implique une connaissance antérieure de la vérité qu'il énonce.

Quelle que soit la part d'Eudoxe dans ces analyses et synthèses conservées, son usage d'analyses est assez constaté pour nous expliquer l'intérêt qu'il a pris aux »trois nouvelles analogies« que — selon les témoignages concordants de Proclus et quelques autres auteurs — il aurait inventées ou »ajoutées aux trois anciennes«. Les nouvelles analogies sont celles que nous exprimons à présent par

$$\frac{a-b}{b-c} = \frac{c}{a}, \quad \frac{a-b}{b-c} = \frac{c}{b}, \quad \frac{a-b}{b-c} = \frac{b}{c}.$$

La question arithmétique qui se soulève ici est de trouver des nombres qui satisfassent à ces équations, ce qui se fait algébriquement en les résolvant par rapport à l'une des trois quantités, les deux autres étant données; il s'agit de déterminer celles-ci de manière à ce que les solutions deviennent rationnelles; les solutions entières en résultent alors sans difficulté, les équations étant homogènes.

A la résolution des dites questions algébriques se prête

immédiatement l'analyse géométrico-algébrique que nous venons de rencontrer dans les cinq propositions d'Euclide XIII. Il faut commencer par dessiner les figures exprimant les trois proportions, soit les équations
$a(a-b) = c(b-c)$, $b(a-b) = c(b-c)$, $c(a-b) = b(b-c)$.
Comme on ne connaît pas encore les valeurs des trois quantités, il ne s'agit pas, toutefois, d'une construction exacte, mais seulement de se procurer des dessins auxquels on puisse attacher l'énoncé des équations.

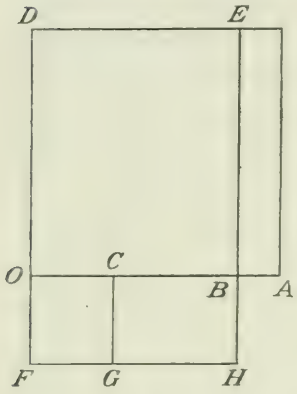


Fig. 1.

On exprimera par exemple la première des trois équations en posant (fig. 1) $OA = a$, $OB = b$, $OC = c$ et en dessinant le carré AD sur OA, et le carré CF sur OC et en prolongeant la droite FG jusqu'à l'intersection H avec la droite EBH perpendiculaire à OA. L'équation demandée s'exprime alors par l'égalité des rectangles AE et CH. La figure montre immédiatement que le rectangle DH est alors égal à la somme des deux carrés, ou bien que

$$b(a+c) = a^2+c^2,$$

ce qui donne une expression rationnelle de b.

Les deux autres équations étant du premier degré par rapport à a, on en peut aussi déduire les solutions rationnelles demandées; mais il est très probable qu'Eudoxe, lorsqu'il a trouvé les questions dignes qu'on s'en occupe, a résolu aussi les autres questions relatives à la détermination d'un terme, les deux autres étant donnés. Alors l'artifice, déjà bien connu, servant à la solution d'une équation du second degré, a demandé l'introduction de certains

rectangles et du carré ayant pour côté la moitié du coefficient de la première puissance de l'inconnue.

Nous ne savons pas jusqu'où Eudoxe est allé dans l'application de ces procédés. Plus tard, on a appliqué, au même effet, les transformations des proportions exprimant les analogies; c'est ce que nous raconte Pappus[1]: on s'en servait pour déduire des proportions données des proportions continues, grâce auxquelles on exprimait les termes d'une analogie comme fonctions linéaires et homogènes de trois nombres, t^2, tu, u^2, qui forment une telle proportion. Le procédé appliqué aurait été, lui aussi, à la disposition d'Eudoxe, créateur de la théorie générale des proportions; mais si les déterminations indiquées remontaient à lui, Pappus l'aurait dit probablement.

Depuis Euclide l'analyse géométrico-algébrique dont nous parlons ici devait prendre une autre forme. Après lui, il suffisait, en effet, de réduire les nouveaux problèmes à des problèmes résolus et à des théorèmes démontrés dans ses Éléments, et cette réduction était facilitée par ses *Data* dont les propositions, sans contenir des vérités nouvelles, étaient adaptées à cet effet. On n'avait plus besoin, par exemple, de réduire des questions dépendant d'équations du second degré à l'usage direct du gnomon. Cette réduction était faite une fois pour toutes dans ses énoncés des propositions 27—29 du livre VI des Éléments, concernant les paraboles de surfaces. Il est vrai que cela se fait sous une forme généralisée que n'aurait pas démandée l'usage pratique; mais le fait qu'Euclide se contente de réduire à de simples paraboles de surface les nombreuses solutions

[1]) Voir la III^e et la IV^e partie de P. Tannery: L'Arithmétique des Grecs dans Pappus (Mémoires de Bordeaux 1880 t. III ou Mémoires scientifiques Vol. I p. 90—98). Voir aussi Pappus ed. Hultsch I p. 84—105.

d'équations du second degré dont il a besoin dans son livre X, montre bien que l'usage de ces paraboles lui était assez familier. C'est par exception qu'il ne s'y contente pas au commencement du livre XIII, exception qui s'explique si les démonstrations qu'il y donne sont empruntées à Eudoxe. De même, dans l'intéressante solution des équations

$$xy = a, \quad y^2 - mx^2 = b$$

que contient la proposition 86 des *Data* (voir mon mémoire sur la Réforme p. 115), il évite un renvoi direct aux paraboles de surfaces ou, ce qui revient au même, aux propositions précédentes (84 et 85) des *Data*. Toutefois il ne fait aucun usage direct de la figure de gnomon, mais se contente de renvoyer à d'autres propositions précédentes des *Data*, dont l'utilité est précisément démontrée par la proposition considérée.

Nous possédons d'ailleurs de l'époque des premiers Alexandrins un excellent spécimen de la forme que prenait l'analyse algébrique après Euclide, et du soin avec lequel on observait dans les œuvres classiques le détail de cette forme. Je veux parler des deux livres d'Apollonius sur la *section de raison.*[1] Il est vrai que nous ne connaissons cet ouvrage que d'après un texte arabe; mais il paraît probable qu'il y est rendu assez fidèlement et sans additions essentielles; car déjà le fait qu'Apollonius consacre deux livres à un seul problème indique qu'il l'avait traité d'une manière assez circonstanciée. Du reste, l'édition conservée répond bien à la mention que fait Pappus de l'ouvrage original.

Le problème dont il s'agit demande la construction d'une droite passant par un point donné H d'un plan et telle que les segments interceptés entre ses points d'inter-

[1] Halley en a publié une version en latin d'après le texte conservé en arabe (1706).

section avec deux droites AB et CD et deux points fixes des mêmes droites soient entre eux dans un rapport donné; mais il ne faut pas croire qu'Apollonius en cherche immédiatement une élégante construction géométrique (comme le faisait naguère Jul. Petersen dans ses »Méthodes et Théories«): sa solution consiste dans une détermination algébrique des segments en question.

Dans le premier livre, il traite, à côté de cas encore plus simples, celui où l'un des points fixes E est le point d'intersection des deux droites données; c'est là, en effet, un cas auquel il est facile de réduire le cas général, qu'il traite dans le second livre. Les cas sont ultérieurement divisés d'après la situation du point donné et les positions des points d'intersection cherchés sur les droites, car on ne savait pas alors distinguer ces positions entre elles par les signes des segments. Comme spécimen des analyses et synthèses nous communiquerons ici celles qui sont relatives au cas II du »lieu« VI du premier livre (pp. 31—37 de l'édition de Halley). Ce qui caractérise ce cas est, si (fig. 2) HT est parallèle à AB et rencontre CD en T et si le point Z est un point fixe de ET, qu'on cherche les droites HKL rencontrant CD en un point K de ZC et AB en un point L de EB, tels

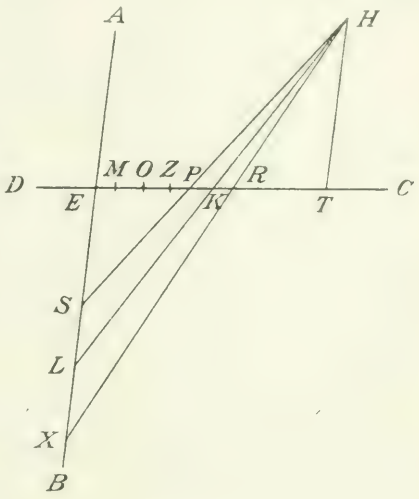


Fig. 2.

que le rapport $\frac{LE}{ZK}$ ait une valeur donnée. Après cette »ecthèse« Apollonius continue l'analyse du problème, supposé résolu dans la figure, de la manière suivante[1]:

»Soit $\frac{TH}{ZM}$ égal au rapport donné $\frac{LE}{ZK}$; alors, le segment TH étant donné, ZM sera également donné comme grandeur et comme position, et le point Z étant donné, le point M sera aussi donné, et comme T est donné, aussi TM sera donné comme grandeur et comme position. Or $\frac{LE}{ZK} = \frac{TH}{ZM}$, ou $\frac{LE}{TH} = \frac{KZ}{ZM}$, mais $\frac{LE}{TH} = \frac{EK}{KT}$; donc $\frac{EK}{KT} = \frac{KZ}{ZM}$ et, par conséquent, $\frac{ET}{TK} = \frac{KM}{MZ}$ et le rectangle $ET \cdot MZ = TK \cdot KM$. Mais $ET \cdot MZ$ est donné, parce que chacun de ses côtés est donné; donc le rectangle $MK \cdot KT$ est donné. Par conséquent, on trouve le point K par l'application de la »parabole en défaut« de ce rectangle sur la droite TM, et comme le point H est donné, la droite HKL sera donnée comme position.«

Cependant l'analyse n'est pas finie par cette réduction (ἀπαγωγή) du problème cherché à une »parabole«, c'est à dire à une équation du second degré. On a prouvé seulement que le point cherché K, s'il existe, peut être déterminé ainsi. Mais avant de commencer la synthèse, il faut encore 1° savoir si les points K déterminés ainsi se trouvent sur le segment ZT utile pour la solution et 2° s'assurer que la condition de possibilité pour la parabole en défaut à laquelle on a réduit le problème proposé est satisfaite. Voilà ce que fait encore Apollonius dans la partie de

[1] Nous nous servons de quelques signes modernes pour abréger l'exposé verbal d'Apollonius, et nous nous dispensons de citer expressément, comme il le fait, les différentes règles connues pour la transformation des proportions.

l'analyse qu'on a désignée par le nom assez peu significatif de »Révolution.«[1]

»Comme dans la synthèse il faut faire le rapport $\frac{TH}{ZM}$ égal à la raison donnée et appliquer, par une parabole en défaut, $TE \cdot ZM$ sur la droite TM, le point K tombera sur TZ« [car s'il tombait sur ZM, on aurait $KM < ZM$ et $KT < TE$ et par conséquent $KM \cdot KT < TE \cdot ZM$]. »Mais il n'est pas toujours possible de mener la droite cherchée; car si $TE \cdot ZM > \left(\frac{TM}{2}\right)^2$ la parabole en ellipse est impossible« [en effet, selon Euclide VI, 27 $TK \cdot KM$ prendra sa plus grande valeur si K est le milieu de TM].

Apollonius remarque encore que si $TE \cdot ZM = \left(\frac{TM}{2}\right)^2$, la parabole se fera d'une manière singulière. Alors le point cherché K sera le milieu de MT, et on aura $TE \cdot ZM = TK \cdot KM$. »Alors $\frac{ET}{TK} = \frac{KM}{MZ}$ et $\frac{TE}{EK} = \frac{MK}{KZ}$ et, comme $KM = KT$, $\frac{TE}{EK} = \frac{TK}{KZ}$, ou bien $\frac{ET}{TK} = \frac{EK}{KZ}$, et ensuite $\frac{ET}{EK} = \frac{EK}{EZ}$. Le segment EK sera donc la moyenne proportionnelle entre ET et EZ. Or ET et EZ et le point E sont donnés. EK et le point K seront donc donnés eux aussi, et, de même, le point M, parce que $KM = TK$«.

Apollonius fait ensuite la synthèse de cette solution singulière. Avant de passer à celle du cas général, il doit comparer le rapport $\frac{LE}{ZK}$ des segments interceptés sur AB et CD par la droite HKL qu'on a trouvée dans ce cas singulier (pour lequel Apollonius réserve ensuite les notations K, L, M) avec ceux qu'y déterminent d'autres droites passant par H. Il démontre au moyen des propriétés, déjà rappelées,

[1] On trouvera l'énumération des différentes parties dont les anciens composaient l'analyse et la synthèse dans Hankel: Geschichte der Mathematik im Altertum und Mittelalter, et dans mes Leçons sur la même matière.

relatives à la parabole en défaut que le rapport $\frac{LE}{ZK}$ a la valeur minimum, et que les rapports augmentent à mesure que la droite s'éloigne de la position singulière HK.

Après ces préliminaires — dont il est superflu de communiquer ici le détail — Apollonius peut donner la synthèse suivante du problème proposé:

»Soit HT parallèle à AB et EK la moyenne proportionnelle entre TE et EZ: menons la droite HK, qui (prolongée) rencontre AB en L. Le rapport donné $\frac{\nu}{\xi}$ peut être $> \frac{LE}{ZK}$. Si $\frac{\nu}{\xi} = \frac{LE}{ZK}$, la droite LH satisfera le problème, et il n'y aura qu'elle qui puisse le satisfaire. Si $\frac{\nu}{\xi} < \frac{LE}{ZK}$, le problème n'est pas possible. Mais si $\frac{\nu}{\xi} > \frac{LE}{ZK}$, on fera $KM = TK$; alors on aura [selon ce qui a été prouvé pour la solution singulière] $ET \cdot ZM = MK \cdot KT$, et $\frac{LE}{ZK} = \frac{TH}{ZM}$. Or $\frac{\nu}{\xi} > \frac{LE}{ZK} = \frac{TH}{MZ}$. Si donc nous déterminons O de façon à ce que $\frac{TH}{ZO} = \frac{\nu}{\xi}$, on aura $ZO < ZM$. Mais comme $ET \cdot MZ = MK \cdot KT$ et $MO \cdot ET > MO \cdot KT$ on voit [par soustraction] que le rectangle $OZ \cdot ET < OK \cdot KT$. Il est donc possible d'appliquer ce rectangle sur OT (Parabole en défaut). En exécutant cette parabole des deux côtés de K, on trouve les points P et R. On mène ensuite les droites HP et HR, qui rencontrent AB en S et X. Je dis qu'alors les deux droites HS et HX satisferont le problème, ou bien que $\frac{\nu}{\xi} = \frac{SE}{ZP} = \frac{XE}{ZR}$. En effet, comme $TE \cdot ZO = OP \cdot PT$, on aura $\frac{ET}{TP} = \frac{PO}{OZ}$, et ensuite $\frac{EP}{PT} = \frac{PZ}{ZO}$. Mais $\frac{EP}{PT} = \frac{SE}{TH}$, et par conséquent $\frac{SE}{TH} = \frac{PZ}{ZO}$, ou bien $\frac{SE}{PZ} = \frac{TH}{ZO} = \frac{\nu}{\xi}$. Chacune des droites HS et HX résout donc le problème, et il

est clair que plus les droites sont près de *HL*, plus sont petits les rapports qui y correspondent«.

Remarquons d'abord qu'Apollonius connaît parfaitement la duplicité des solutions d'une équation du second degré, bien entendu dans les cas où toutes les deux racines sont positives; car il ne connaît pas plus les solutions négatives qu'imaginaires. Nous avons donc ici un démenti net du préjugé qui veut que les Grecs aient ignoré, ou du moins négligé cette duplicité[1], et qu'en général ils se soient souciés fort peu du nombre des solutions d'un problème. Au contraire, l'importance qu'attachait Apollonius au nombre des solutions possibles, ressort du résumé détaillé, qu'à la fin de son livre I, il fait des nombres de solutions qu'aura son problème pour les différentes positions des droites et points donnés (p. 58—59). Ce résumé embrasse à la fois les solutions, qui pour nous répondraient à des signes différents du rapport donné.

On s'explique du reste très bien pourquoi on ne trouve rien chez Euclide sur la duplicité des racines d'une équation du second degré ou d'une parabole en défaut. Abstraction faite des généralisations qu'on lui doit, il s'agit en effet pour lui comme pour ses prédécesseurs de trouver un rectangle, en connaissant la surface et la somme des côtés, ou un produit, en connaissant la valeur et la somme des facteurs, et ce problème n'a en réalité qu'une seule solution, tant qu'on ne fait aucune distinction entre les deux facteurs ou côtés. Le but visé par Apollonius en introduisant sa parabole, est au contraire de déterminer au moyen d'elle un point d'un segment *OT* dont les deux bouts, *O* et *T*, jouent des rôles différents. Alors on aura des

[1] Paul Tannery réfute le même préjugé par le renvoi à d'autres exemples (Mémoires scientifiques I p. 93).

solutions différentes suivant qu'on fera la section de OT dans l'un ou l'autre des deux points qui décomposent OT en deux segments qui font un rectangle de surface donnée.

On voit du reste combien la subdivision prescrite pour l'analyse et la synthèse est bien faite pour retenir l'attention sur tout ce que demande la discussion complète d'un problème résolu au moyen d'une équation du second degré. La première partie de l'analyse, ἀπαγωγή, contient la réduction à une telle équation. Sa seconde partie n'est pas une anticipation de la première partie de la synthèse, comme le semble indiquer le nom assez impropre de »Résolution« qu'on lui a donné: on y démêle avant tout deux questions qui doivent précéder la synthèse, 1° si par hasard, à côté des conditions exprimées par l'équation à laquelle on a réduit le problème dans la première partie de l'analyse, il en existe d'autres que les solutions cherchées doivent satisfaire en même temps, et, alors, si elles sont satisfaites, elles aussi, par les racines de l'équation, et 2° si la solution de l'équation est possible. De cette façon on découvre comment il faut délimiter les différents cas et les caractériser par des diorismes pour énoncer les problèmes sous une forme qui permette une synthèse utile. Ensuite la première partie de la synthèse contiendra la construction qui fera paraître une multiplicité éventuelle des solutions; sa seconde partie contient la démonstration de la justesse de ces solutions.

Nous avons parlé ici d'équations du second degré parce que les constructions prescrites par Apollonius poursuivent les mêmes buts qu'elles, en se servant seulement d'une forme géométrique pour assurer aux résultats la généralité que comportait la manière dont on avait établi les principes géométriques. Dans cette forme, il ne s'agissait pas

de déterminer une quantité inconnue, mais un point ou une »section« [1] d'une droite donnée. Au lieu de proposer une équation à laquelle l'inconnue doive satisfaire, on regardait, dans l'analyse, comme concédée une figure où la section satisfasse aux conditions demandées. La forme géométrique présentait à cet égard une plus grande généralité, parce qu'elle ne demande rien de particulier à l'origine du segment qu'on regarde de nos jours comme l'inconnu, de même que l'usage perpétuel de rapports dispense d'un choix particulier de l'unité. Nous avons vu déjà que la première de ces circonstances a permis de représenter par une seule figure, celle du gnomon, les mêmes choses qu'actuellement on exprime par trois formules, ou bien par deux formules en attribuant des signes aux quantités représentées par une lettre. On aura remarqué aussi que le besoin de signes n'existe pas dans les opérations purement géométriques, où toutefois les figures employées se modifient suivant les situations relatives des différents points soumis à ces opérations. L'opuscule d'Apollonius qui nous occupe montre du reste que cette dernière circonstance a demandé une décomposition assez pénible d'une recherche générale en cas particuliers, décomposition qu'on évite à présent en attribuant des signes aux quantités représentées par les symboles algébriques. Cet inconvénient n'empêche pas d'ailleurs

[1] Ce qui pour nous est la détermination de la racine positive de l'équation $a^2 - ax = x^2$ était pour les Grecs la section du segment a en moyenne et extrême raison. Le titre Περὶ τῆς τομῆς d'un travail d'Eudoxe, dans lequel »la section« n'est pas définie ultérieurement ne pourrait-il pas désigner un traité général d'équations du second degré sous forme de sections? Nous n'avons aucun moyen de répondre à cette question; mais, comme nous l'avons vu, les recherches algébriques d'Eudoxe ont pu en tout cas servir de paradigmes de recherches de cette nature. Nous renvoyons aussi aux titres de trois opuscules d'Apollonius qui indiquent des études de »sections« (τομή ou ἀποτομή).

Apollonius de faire paraître l'uniformité des procédés qu'il applique à des cas différents: il la souligne par l'emploi des mêmes expressions lorsqu'il applique la même opération ou le même raisonnement à des figures différentes.

L'opuscule d'Apollonius a donc été bien fait pour apprendre à ses lecteurs et élèves de bien appliquer l'algèbre géométrique à la solution et discussion d'un problème général: elle permet à la fois de montrer la différence des aspects que prendra le problème et la solution dans les différents cas, et de faire paraître la généralité des vues. Un tel usage a demandé l'étude détaillée de divers exemples pratiques, de même que l'habitude de l'analyse algébrique moderne n'est acquise qu'au moyen d'exercises de calcul algébrique. D'après Pappus[1], d'autre opuscules actuellement perdus d'Apollonius ont dû servir à propager cet emploi de la solution d'équations du second degré. Le thème des deux livres sur la *section de surface* n'a différé de celui dont nous venons de parler que par la circonstance qu'on y donne le rectangle au lieu du rapport, formé par les deux segments interceptés sur les droites données. A cause de cette analogie, Apollonius a pu y appliquer, comme nous le rapporte Pappus, la même division en »lieux« et »cas«, ce qui a permis à Halley de restituer cet ouvrage d'une manière dont la correction quant au fond ne saurait être mise en doute; seulement Halley obtient une plus grande brièveté en réunissant dans des exposés plus généraux des cas qu'Apollonius a traité séparément.

Dans les deux livres sur la *section déterminée*, Apollonius s'est occupé du problème de trouver sur une droite donnée contenant quatre points donnés A, A', B, B', un point M tel que le rapport de deux rectangles $\frac{AM \cdot A'M}{BM \cdot B'M}$

[1] Édition de Hultsch pp. 640—645, 670—673.

ait une valeur donnée. Ne s'occupant que des segments d'une droite, ce problème est déjà algébrique et conduira en général à une équation du second degré. Les deux points M qu'elle sert à déterminer, si nous attribuons à la valeur du rapport un signe + ou —, feront une couple de l'involution (comme nous disons aujourd'hui) déterminée par les couples AA' et BB'. Ne connaissant pas les signes, Apollonius doit avoir trouvé deux de ces couples si les racines qui y correspondent sont réelles, et l'épreuve de cette réalité (ou possibilité) doit avoir constitué un point principal de sa recherche, qui a contenu — selon Pappus, et correspondant aux différéntes dispositions des points donnés — 8 indications du maximum ou du minimum du rapport donné, c'est-à-dire de valeurs de ce rapport auxquelles correspondent des points doubles de l'involution. Parmi les cas traités ont aussi figuré, selon Pappus, ceux où A et A', ou B et B' coïncident en un point double. Les lemmes que Pappus a trouvé utile d'ajouter [1] montrent qu'Apollonius n'a pas négligé les cas où le rapport donné est égal à un, et où l'équation exprimant la relation proposée sera du premier degré et déterminera le point que nous appelons aujourd'hui le point central de l'involution. Les propriétés exceptionnelles de ce point ont pu, dans l'ouvrage d'Apollonius, ou donner lieu à une division ultérieure en »cas« ou trouver place dans la seconde partie (la »résolution«) de l'analyse des »cas« déjà introduits, à côté des conditions de possibilité.

Quoi qu'il en soit, l'opuscule doit avoir contenu, sous la forme qu'on donnait alors aux équations du second degré, une discussion étendue de la question algébrique que nous exprimerions de nos jours par l'équation

[1] Édition de Hultsch pp. 704—771.

$$x^2 + ax + b + \lambda(x^2 + a_1 x + b_1) = 0,$$

et l'identité effective des problèmes doit avoir amené un parallélisme assez complet des traitements sous différentes formes.

Encore un opuscule d'Apollonius doit avoir contenu, sous la même forme géométrique, des solutions d'équations du second degré. Dans ses deux livres sur *les intercalations* (νεύσεις), il s'occupe du problème de mener par un point donné une droite telle que deux lignes données interceptent sur elle un segment de grandeur donnée. Il s'agissait là de substituer à l'opération mécanique, qui pouvait être bonne pour la pratique, des procédés qui, par le seul usage de lignes bien définies géométriquement, pussent fournir la même rigueur que présentent nos solutions algébriques. Avant tout, on devait connaître les moyens d'exécuter l'intercalation, dans les cas où c'est possible, au moyen de droites et de cercles, ce qui équivaut à une solution par une équation du second degré. Tel est le but d'Apollonius, et par conséquent il a du réduire ces problèmes à des paraboles de surface, qui remplacent chez lui les équations du second degré.

Pappus nous fait connaître les problèmes de cette nature, dont s'occupe Apollonius. Dans l'un de ces problèmes, les lignes données sont des cercles, et le point donné, un de leurs points d'intersection avec la droite passant par leurs centres; il y en a d'autres qui sont, du moins à nos yeux, des cas particuliers du même problème. Au moyen des procédés dont il rend compte dans son ovrage sur *les lieux plans* — opuscule qui précède dans l'énumération de Pappus celui dont nous nous occupons ici — Apollonius a pu réduire ce problème à une équation du second degré.

Dans un autre, le point donné est un des sommets

d'un losange tandis que les lignes données sont les côtés qui enferment le sommet opposé. C'est un problème qui a été suggestif pour les mathématiciens de la Renaissance. Tandis qu'il fournit à GIRARD un exemple de l'usage des signes, DESCARTES s'en sert pour illustrer les équations du quatrième degré qui, au moyen de la solution d'une équation du second degré, se décomposent en deux équations du second degré; mais NEWTON a montré que, par un choix heureux de l'inconnue, on peut parvenir immédiatement à une équation de ce degré.[1] Nous possédons de l'antiquité une élégante construction attribuée à un certain HÉRACLITE[2]; mais cette attribution nous indique que son élégance, qui est de nature purement géométrique, n'est pas due à APOLLONIUS. Il est donc probable que dans son ouvrage, destiné à montrer une méthode pour traiter d'une manière générale les problèmes de cette nature, la solution a été obtenue par des considérations moins adaptées à la question géométrique particulière, ou bien par des procédés algébriques; mais alors APOLLONIUS aura su surmonter des difficultés algébriques assez sérieuses, soit que sa solution, traduite dans le langage de l'algèbre moderne eût surtout ressemblé à la solution de DESCARTES, soit qu'elle se fût plutôt rapprochée de celle de NEWTON.

Tout en admettant que, dans les travaux mentionnés ici, on obtient la même chose que nous réalisons actuellement par la détermination d'une quantité cherchée au moyen d'une équation du second degré, et en reconnaissant le parallélisme des procédés exécutés sous des formes diffé-

[1] Voir ma »Geschichte der Mathematik im XVI and XVII Jahrhundert p. 212—214 (Édition danoise p. 299—301).

[2] Kegelschnitte im Altertum p. 281 (Édition danoise p. 182).

rentes, on insistera peut-être encore sur la grande différence que ces formes présentent entre elles, l'une en déterminant une »section« cherchée, l'autre en déterminant une quantité cherchée; mais comme il résulte de ce que j'ai dit plus haut, cette différence dépend plutôt de la manière dont on a essayé d'obtenir une rigueur satisfaisante dans l'application des procédés à des questions générales ou plus abstraites. Après Platon, on cherchait à assurer cette rigueur par les formes géométriques soumises aux suppositions bien formulées de la géométrie; mais ce qui avait intéressé d'abord, et ce dont on avait besoin toujours dans les problèmes concrets, c'était la détermination de quantités inconnues. Pour construire un pentagone régulier, on avait besoin du rapport de son côté à une diagonale; on trouvait que ce rapport serait celui du plus grand segment d'une droite divisée en extrême et moyenne raison à la droite elle-même, ou celui des deux segments entre eux. C'est la détermination de cette quantité inconnue qu'on réduisait à une parabole en excès. Héron sait très bien dans ses *Metrica* appliquer les paraboles à la détermination de quantités, et ayant pour but d'illustrer par des exemples numériques les propositions générales des anciens, il fait probablement à cet égard ce qu'on avait déjà eu en vue en établissant les propositions générales. Il n'y a guère que Diophante qui nous apprend comment les Grecs ont traité des questions concrètes et numériques sans les ramener à des propositions abstraites et générales — seules ses dénominations géométriques rapellent encore la forme géométrique de ces propositions. C'est ainsi qu'il nous apprend directement la formation d'équations à une seule inconnue, et grâce aux notations mentionnées à la page 1 il sait représenter ces équations sans faire usage des formes géomé-

triques ni des proportions continues (voir p. 24). Malheureusement nous ignorons jusqu'à quelle époque remonte une telle formation directe des équations; mais l'étendue et la grande variété des questions algébriques auxquelles elle se trouve appliquée dans l'ouvrage de DIOPHANTE montre qu'il ne se s'agit pas d'un progrès individuel, mais bien des fruits d'un long développement. Je ne crois pas que celui-ci a eu lieu en opposition avec la géométrie mais plutôt sous l'abri des anciennes démonstrations géométriques d'un caractère général. C'était à elles qu'il fallait avoir recours pour assurer, par exemple, la justesse et la généralité de la solution des équations du second degré; on s'en est servi à peu près comme nous employons aujourd'hui les formules algébriques. En même temps qu'on assurait ainsi les opérations dont on se servait pratiquement, on faisait, successivement, de nouvelles observations et inventait de nouveaux artifices, qu'après le temps des grands géomètres on ne savait plus condenser dans des théorèmes et problèmes généraux, bien formulés et exactement démontrés, susceptibles en un mot d'être exposés dans des ouvrages que la postérité aurait pu conserver et commenter à côté de ceux des premiers géomètres alexandrins. C'est pour cette raison que du long intervalle entre EUCLIDE et DIOPHANTE le nombre est singulièrement restreint des ouvrages conservés qui nous semblent préparer directement les recherches dont DIOPHANTE nous fait connaître les fruits et qui puissent servir à illustrer la connexion entre la géométrie scientifique et les progrès des procédés pratiques. Déjà CANTOR et surtout P. TANNERY ont attiré l'attention sur ces ébauches éparses des doctrines de DIOPHANTE. Je me bornerai donc à en citer deux exemples pour y

signaler la connexion que cette préparation a eu avec la géométrie grecque et l'algèbre qui s'y était rattachée.

Comme premier exemple je citerai la détermination, déjà mentionnée à la page 39, de la valeur des termes des différentes médiétés, c'est-à-dire de la formation de certaines équations du second degré à racines entières. Il s'agit ici de la solution générale de questions du même ordre que les questions plus concrètes dont Diophante s'occupe, en se contentant toutefois d'obtenir des solutions rationnelles.

Notre second exemple est tiré de la collection de questions jointe aux éditions des ouvrages de Héron et conservée dans le livre dit *geponicus*, et dont M. Heiberg a publié un texte plus exacte.[1] Les questions indéterminées résolues dans cette collection sont de la même nature que celles dont s'occupe Diophante; leur invention et solution ont demandé une habitude assez développée de manipulations algébriques; mais, de même que le fait ordinairement Diophante, on se contente d'en trouver des solutions, sans s'assurer que les solutions trouvées sont les plus générales. Nous y signalerons pourtant un emprunt fait à la théorie générale d'Euclide. Dans la solution d'un de ces problèmes indéterminés, on invoque le fait que la surface d'un triangle rectangle dont les côtés sont exprimés par des nombres entiers, doit être divisible par 6. La connaissance de ce fait a du être tirée de la solution générale de l'équation indéterminée $x^2+y^2=z^2$, qu'on trouve dans Euclide X, 28 Lemme 1, et que nous avons rappelée à la page 18. Cette solution conduit à l'expression suivante de la surface aux côtés x, y, z:

$$\frac{1}{4}c^2mn(m-n)(m+n),$$

[1] Voir Heiberg und Zeuthen: Einige griechische Aufgaben der unbestimmten Analytik (Bibliotheca mathematica 8^3).

dont on remarque sans difficulté la divisibilité par 2 et 3 et par conséquent par 6.

IV. Approches de la géométrie analytique; théorie des sections coniques.

Le plus grand témoignage que nous ayons de l'importance de l'algèbre grecque, de son état élevé de développement et de sa portée scientifique, c'est la théorie des sections coniques que nous révèlent déjà les applications, qu'Archimède fait de ces courbes, et dont nous apprenons à connaître l'enchaînement complet à travers la grande œuvre d'Apollonius. Je m'en suis occupé dans un traité, publié en 1885 en danois dans les Mémoires de l'Académie Royale des Sciences et des Lettres de Danemark (t. III de la 6e série, section des sciences) et dont ensuite une traduction en allemand a paru en 1886 (Die Lehre von den Kegelschnitten im Altertum, Kopenhagen, A. F. Høst & Søn).

Le but que je me suis proposé dans ce travail était de retrouver les méthodes grâce auxquelles les anciens Grecs avaient pu parvenir à des résultats que la géométrie moderne n'a reconquis qu'assez lentement. A cet effet, il fallait fixer l'attention sur les idées directrices de démonstrations qui semblent souvent à un mathématicien moderne assez pénibles à suivre, mais qui doivent avoir été plus familières aux lecteurs anciens d'Apollonius; sinon il n'en aurait pas trouvé. La nature géométrique des démonstrations, qui leur donne une forme si différente des considérations algébriques modernes, pouvait porter à en chercher une clé géométrique en se demandant si les anciens n'auraient pas eu recours aux considérations stéréométriques, fondées sur la théorie de la perspective que, dans les temps modernes, Desargues, Pascal et, plus tard, Poncelet ont utilisées

avec tant de succès. Il n'en est rien: dans l'antiquité, la recherche stéréométrique des propriétés des coniques se bornait essentiellement au strict nécessaire et l'usage qu'on en faisait jouait un rôle assez inférieur à celui des considérations planimétriques. L'idée de faire des propriétés des foyers le point de départ d'une étude géométrique de ces courbes, comme l'ont éssayé Steiner et Zeuthen dans le siècle passé, n'a pas non plus été utilisée par les anciens auteurs. Faut-il donc croire que nous ayons perdu complètement la connaissance de voies qui avaient conduit autrefois à la découverte de résultats si importants et généraux, voies que nous devrions alors nous empresser de retrouver? Non. L'algèbre moderne et la géométrie analytique, qui nous servent ajourd'hui à établir les mêmes résultats sont dues elles-mêmes à un développement ultérieur et à une transformation importante des procédés antiques, transformation due à la substitution des symboles algébriques modernes aux figures géométriques qui servaient dans l'antiquité à marquer des opérations semblables. C'est pourquoi un mathématicien moderne peu accoutumé à la forme géométrique de l'ancienne algèbre saisira mieux l'idée directrice d'une démonstration d'Apollonius au moyen d'une traduction dans la langue algébrique qui lui est plus familière; après quoi il comprendra aussi la manière dont il fallait exprimer alors une démonstration analogue.

C'est par une analyse de la théorie grecque des sections coniques que je suis conduit à voir comment la géométrie grecque sert d'organe à une algèbre assez développée pour fournir les moyens de construire une théorie aussi élevée, théorie qui fournit de son côté un développement ultérieur de l'organe qui avait servi à la construire. De cette algèbre géométrique, dont la grande portée m'était

apparue dans ses applications à la théorie des coniques, je devais chercher les fondements et les applications antérieures dans les Éléments d'Euclide et ailleurs. C'est cette analyse historique que j'ai essayé de reporter ici le plus loin possible afin de construire dans le présent travail une synthèse du développement successif de l'algèbre grecque depuis ses premières origines jusqu'à ce qu'elle ait atteint, dans la théorie des coniques, sa culmination. Comme d'ailleurs on trouvera la plupart des observations que je pourrais faire au sujet de cette théorie grecque dans mon ouvrage déjà cité, je me bornerai à présent à en tirer assez pour faire ressortir le rapport de liaison qui la rattache au développement algébrique antérieur dont je viens de m'occuper ici.

Les procédés suivis dans le traité des coniques d'Apollonius ont été préparés par les opuscules du même auteur dont nous avons déjà parlé. L'usage direct des différentes »sections« dont il s'occupe dans trois d'entre eux, se rapporte certainement à des propositions particulières du livre III, et j'en ai parlé dans les chapitres XV et IX de mon livre cité, mais, à côté de cela, la discussion très détaillée que dans ces opuscules Apollonius entreprend de problèmes généraux qui dépendent d'équations du second degré a dû préparer ses élèves à ne rien oublier de ce que demande une telle discussion. L'uniformité de son exposé des différents cas leur aura donné une idée des principes généraux qu'on poursuit partout, et ils auront été ainsi mis en état de bien saisir la généralité des points de vue qu'il prend ensuite dans son traité des coniques, et que lui permet la considération, inventée par lui, des »deux hyperboles opposées« — c'est-à-dire des deux branches d'une hyperbole — comme parties d'une seule courbe. Cette généralité se

trouve exprimée de façon formelle quand il énonce, comme il le fait souvent, des théorèmes concernant »une section conique, ou une circonférence de cercle, ou des sections opposées«, c'est-à-dire ce que nous appelons aujourd'hui une conique quelconque. Toutefois, à défaut de la notion moderne d'un »point à l'infini« il ne lui a pas été possible d'obtenir toujours une telle généralité formelle: mais alors les lecteurs des dits opuscules auront été habitués à n'oublier aucun des cas différents qu'il fallait considérer à part. Ils auront été accoutumés à voir la nécessité de délimitations (»diorismes«) exactes et à remarquer l'importance des maxima et minima qui en résultent, ainsi que l'utilité d'une indication des nombres de solutions qu'on trouve dans les différents cas, et à comprendre ainsi certains avantages des théorèmes généraux. Apollonius signale plusieurs fois dans ses préfaces aux divers livres de ses Coniques ces avantages, qui se rattachent non seulement à la solution des problèmes dits plans (qui dépendent d'équations du second degré) mais aussi aux solutions des problèmes, appelés solides, que l'algèbre moderne fait dépendre d'équations du 3e (ou 4e) degré, et dont les anciens obtenaient au moyen des coniques une solution sous forme géométrique.

D'une manière plus directe que dans les dits opuscules, Apollonius a préparé dans ses deux livres sur *les lieux plans* les procédés dont il fait usage dans son traité des sections coniques. Les Grecs appellent »plan« tout lieu géométrique qui est une droite ou un cercle. Dans le dit opuscule, dont nous connaissons le contenu par la mention qu'en fait Pappus et par les lemmes qui doivent selon le même auteur servir d'appui aux démonstrations qui s'y

trouvent[1], Apollonius s'attache moins à énumérer — comme le fait J. Petersen dans ses Méthodes et Théories — les lieux de cette nature qui sont les plus utiles pour inventer des constructions élégantes au point de vue géométrique, qu'à trouver des critères, de la même nature que ceux que nous fournit à présent la géométrie analytique, pour décider si un lieu géométrique est plan.[2] Le théorème I, 6 énonce par exemple, abstraction faite de quelques généralisations qui ne nous intéressent pas à ce point de vue, que le lieu d'un point dont les distances de deux droites données satisfont à une équation du premier degré, est une droite, et la généralisation, contenue dans le théorème I, 7, où les deux droites sont remplacées par un nombre quelconque de droites, montre qu'on a su profiter de la forme linéaire de cette équation.

De même, les lieux traités dans le livre II montrent qu'Apollonius a très bien su profiter de cette propriété d'un cercle que la somme des carrés des projections des rayons sur deux axes orthogonaux a une valeur constante, propriété identique à celle que nous exprimons de nos jours par l'équation du cercle rapporté aux mêmes axes. Grâce à cette propriété d'un cercle quelconque il a su démontrer le théorème assez général (II, 5) qui énonce que le lieu d'un point M tel que la somme des surfaces de figures semblables à des figures données, et construites sur les droites joignant M à un certain nombre de points fixes, aient une valeur donnée, est un cercle.

C'est l'essai de démontrer les théorèmes généraux que nous avons rappelés ici qui a suggéré à Fermat l'idée de

[1] Pappus ed. Hultsch pp. 660—671 et 852—865.

[2] Voir le chap. X de ma Théorie des coniques dans l'antiquité, et, pour les renvois à Fermat, ma Note sur les coordonnées dans l'antiquité (Bulletin de l'Académie de Danemark (1888)).

faire usage de coordonnées orthogonales, invention qu'il faisait indépendemment de Descartes. Guidé par la mention que font Apollonius et Pappus des »lieux à trois ou quatre droites«, il appliquait ensuite les coordonnées à l'étude des sections coniques. Nous y voyons un signe qu'en réalité, malgré la grande différence des symboles algébriques, Apollonius de son côté a poursuivi, dans ses démonstrations des mêmes résultats, dont nous connaissons seulement ses énoncés, des idées semblables à celles que dans la géométrie actuelle on exprime au moyen de l'algèbre moderne.

Dans la géométrie antique, ce qu'on appelait le symptôme d'une courbe jouait le même rôle que, de nos jours, son équation. Seulement, celle-ci s'exprime pas nos symboles algébriques, tandis que le symptôme ancien exprimait une propriété des points de la courbe sous une forme géométrique où, par exemple, les produits de deux facteurs étaient remplacés par des rectangles. L'usage de tels symptômes convenait bien à la manière dont on réduisait la recherche d'une quantité inconnue à celle de l'intersection de deux lieux. Il est probable que d'abord on a eu en vue la construction mécanique qu'on obtient de cette manière: dans les cas des problèmes dits plans, où les constructions peuvent être exécutées au moyen de la règle et du compas, elle a une grande valeur pratique que certainement on a apprécié toujours comme nous le faisons actuellement; mais les constructions mécaniques (ou par un certain nombre de leurs points) d'une quadratrice, d'une spirale d'Archimède, d'une conique, d'une conchoïde ou d'une cissoïde ne sont ni assez simples ni assez exactes pour avoir été destinées à des buts pratiques:[1] c'était en formulant bien

[1] L'usage de la conchoïde remplace sans doute les intercalations (νεύσεις) pratiques dont nous avons parlé, et on nous relate comment

les symptômes de ces courbes, de même que ceux d'une droite ou d'une circonférence de cercle, qu'on avait le moyen de désigner rigoureusement les points d'intersection de deux courbes dont on connaissait les symptômes, ensuite leurs projections sur une droite, et par cela des quantités cherchées, — et qu'on pouvait s'en servir pour discuter les solutions des problèmes analytiques. On réduisait ainsi la détermination d'une racine d'équation, ou quantité qu'aujourd'hui nous ferions dépendre d'une équation à une inconnue, à la détermination de deux symptômes de courbes, ce qui correspond à la détermination de deux équations à deux inconnues. [1]

s'effectuait sa construction instrumentale, mais son étude fournit avant tout les moyens de discuter exactement les résultats. La division d'un arc de cercle et la rectification et la quadrature d'un cercle s'obtiennent plus facilement et aussi exactement pour la pratique par des essais ou par tels moyens mécaniques dont se servaient déjà les Égyptiens que par l'usage d'une quadratrice ou d'une spirale.

[1] C'est un tel emploi théorique qui a donné leur véritable valeur aux solutions obtenues au moyen d'intersection de courbes, — ce que la postérité n'a pas toujours observé.

Après avoir complété la solution d'équations du 3e degré d'une manière qui la fait dépendre numériquement soit de l'extraction de racines cubiques soit de l'usage de tables trigonométriques, Viète fait remarquer que par ces procédés on la réduit aux deux opérations suivantes, dont l'exécution sous forme géométrique avait occupé les anciens, savoir à la détermination de deux moyennes proportionnelles et à la trisection de l'angle. Il est permis de voir dans cette remarque de Viète une conséquence naturelle de la continuité du développement de la mathématique; mais l'exécution géométrique des dites opérations, qui se faisait au moyen de sections coniques, allait perdre son intérêt devant l'usage immédiat des nouvelles découvertes dues aux prédécesseurs italiens de Viète et à Viète lui-même. Nous comprenons aussi que Descartes et ses contemporains aient été contents de joindre à leurs solutions d'équations du 3e et du 4e degré des solutions où, comme les anciens, on fait usage de l'intersection de sections coniques, et d'appliquer la géométrie analytique à inventer des solutions constructives d'équations de degrés supérieurs où l'on fait usage de l'intersection de courbes supérieures. Et pourtant les véritables progrès obtenus par leur nouvelle algèbre

C'est à un tel usage que se rapporte la première mention directe des sections coniques qui nous soit conservée (Voir le chap. XXI de mon livre ci-dessus cité). Déjà Hippocrate de Chios avait réduit le célèbre problème de la duplication (ou multiplication du cube, soit la résolution de l'équation cubique pure), à la détermination de deux moyennes proportionnelles, réduction conforme, du reste, à la représentation exacte et générale des puissances dont se sert Euclide (voir p. 24). C'est ainsi que la racine $x = \sqrt[3]{a^2 b}$ sera déterminée par les proportions

$$a : x = x : y = y : b, \qquad (1)$$

ou bien par les équations

$$x^2 = ay, \quad y^2 = bx, \quad xy = ab. \qquad (2)$$

En représentant ces quantités, à la manière des anciens, au moyen de rectangles et de carrés, on réduira le pro-

tendaient à rendre superflues les représentations géométriques dont s'étaient servis les anciens au même effet.

Moins innocente est une autre conséquence de l'adhésion aux formes antiques. Même là où on poursuit des buts pratiques, on préfère jusqu'à nos jours à toute autre une exécution graphique à laquelle les anciens donnaient la préférence à cause de ses avantages théoriques: à ce point de vue elle se trouvait être la plus rigoureuse, parce qu'elle se fondait sur des définitions et des postulats bien formulés, tandis qu'au point de vue pratique ni l'une ni l'autre exécution mécanique n'est d'une exactitude absolue. En se servant toujours d'intersections de courbes, on a poussé dans nos écoles de dessin technique la finesse des tracés et l'exactitude des constructions, non seulement de droites et de cercles, mais aussi des courbes réunissant des points construits, jusqu'à une excellence qui procure aux constructions exécutées par ces moyens une exactitude satisfaisant même les exigences sévères de la technique de nos jours; mais en présence de ces moyens on a négligé d'habituer en même temps les élèves aux autres moyens techniques qui amèneraient souvent plus facilement une plus grande exactitude: on a laissé aux élèves le soin d'en inventer eux-mêmes, soit pour contrôler les constructions exécutées par les moyens obligatoires, soit pour les substituer à celles-ci par une sorte d'escroquerie. Il était donc temps que M. Hjelmslev rendît compte dans son ouvrage intitulé *Geometriske Experimenter* (Kjobenhavn 1913) des moyens dont le dessinateur technique devra se servir à côté de ceux qui ont été regardés jusqu'à présent comme obligatoires.

blème à la détermination d'un point (x, y) tel que les figures formées par les segments x, y, a et b, où a aussi bien que x est une abscisse, b comme y une ordonnée, aient les relations (2). Chacune d'elle sera le »symptôme« d'une courbe, et deux d'entre elles déterminent par leur intersection le point (x, y) et par suite la quantité cherchée x.

Tout cela est une conséquence de la manière dont les anciens représentaient les produits, et de leur usage de courbes bien définies dont nous venons de parler; mais c'est à Ménechme qu'on doit l'observation que les mêmes courbes sont des sections d'une cône à base circulaire. Cette observation lui a été très utile, au point de vue théorique, dans ses efforts pour réduire autant que possible le nombre des définitions et postulats géométriques (voir mon mémoire sur la Réforme p. 39—46). Elle lui permettait, en effet, de rattacher la véritable existence des courbes (2) à celle des cercles. Il ne s'agissait donc pas pour lui de rechercher, dans un but géométrique, la nature des sections planes d'un cône ou d'un cylindre circulaire — l'ancienne designation de l'ellipse sous le nom de »thyreos« semble du reste indiquer que déjà avant lui on s'était occupé de l'étude géométrique de ces sections —; ce qu'il voulait démontrer c'était que les courbes qui l'intéressaient à cause de leur application algébrique appartiennent bien à cette catégorie de courbes qu'on avait déjà rencontrées dans la stéreotomie. On comprend donc que ni Ménechme ni ceux qui avant Apollonius s'occupaient particulièrement de l'étude des sections coniques ne se soient souciés des propriétés de toutes les sections d'un cône circulaire quelconque: il leur suffisait d'avoir représentées comme sections coniques les courbes dont on faisait un usage algébrique, usage auquel se rattachait alors une étude planimétrique de plus en plus

détaillée des mêmes courbes. Voilà pourquoi on pouvait se contenter de considérer les sections faites à des cônes droits par des plans perpendiculaires à une génératrice. Cette restriction n'a pu être due aux difficultés que pourrait causer l'étude d'autres sections, difficultés qui n'auraient différé en rien de celles qu'on avait déjà surmontées pour connaître les sections particulières dont on se servait pour avoir un point de départ bien fondé des recherches sur les courbes planes introduites en vue de leur usage algébrique. L'introduction au livre d'Archimède sur les conoïdes et sphéroïdes nous apprend en effet les moyens dont on se servait alors pour trouver les propriétés planes des sections, et Archimède, qui en a besoin pour ses propres recherches géométriques, les applique aussi à d'autres sections d'autres cônes (Voir le chapitre II de mon livre cité). C'est donc moins une nouvelle découverte que le désir de se placer à un point de départ aussi général que possible qui à porté Apollonius à commencer son traité général de ces courbes par la considération d'une section plane quelconque d'un cône circulaire quelconque.

En même temps, Apollonius maintenait ou plutôt soulignait la nature algébrique du symptôme planimétrique d'une conique. D'après ce que nous apprend le livre cité d'Archimède, on s'était servi jusqu'alors, pour rapporter une ellipse ou hyperbole à deux diamètres conjugués, de la relation

$$\frac{y^2}{xx'} = \text{constante},$$

où y est l'ordonnée rapportée à ces axes de coordonnées, x et x' les segments interceptés sur l'axe des abscisses entre cette ordonnée et ses intersections avec la courbe. C'est le même symptôme qu'Apollonius a eu l'idée de

rapprocher de la forme antique d'équations du second degré. En effet, si l'on désigne la longueur du diamètre par a et la valeur constante du rapport par $\frac{p}{a}$, on peut énoncer l'équation

$$y^2 = px \mp \frac{p}{a}x^2,$$

correspondant respectivement à l'ellipse, pour laquelle on aura $x' = a - x$, et à la branche d'hyperbole voisine du terme du diamètre que nous avons pris pour origine des x, pour laquelle on aura $x' = a + x$. Alors le carré y^2 doit être égal à un rectangle appliqué sur la droite p de telle sorte que le rectangle en défaut (ellipse) ou en excès (hyperbole) soit semblable à un rectangle aux côtés p et a. Les opérations servant à déterminer les deux quantités x et x' correspondant à une valeur donnée de y seront donc les mêmes qu'Euclide prescrit dans VI, 28 et 29 pour une parabole en ellipse ou en hyperbole. Voilà l'origine des noms que, depuis Apollonius, portent les deux courbes en question. Quant aux hyperboles, à défaut de nos signes modernes, il fallait appliquer l'opération mentionnée à chacune de ses deux branches séparément, ce qui n'empêchait pas de les regarder comme parties d'une seule courbe. La courbe $y^2 = px$ a reçu le nom de parabole parce qu'ici x est le côté du rectangle résultant d'une simple parabole du carré y^2 sur p.

Il faut remarquer qu'indépendamment de l'angle que fait le diamètre avec les ordonnées (demi-cordes) la figure auxiliaire servant à représenter l'équation que nous avons écrite ici sous forme algébrique moderne, se compose de rectangles. Cette indépendance de la figure géométrique dont elle sert à représenter les propriétés algébriques, fait paraître sa nature d'instrument algébrique. Le résultat de

la transformation de coordonnées, grâce à laquelle on rapporte la même courbe de la même manière à un nouveau diamètre, s'exprime par une nouvelle figure orthogonale de la même nature.

Dans mon livre déjà cité, j'expose comment, en partant des symptômes planimétriques des coniques que nous avons décrits ici, Apollonius établit les propriétés de leurs diamètres conjugés, axes, tangentes, asymptotes, foyers, de pôles et polaires, aussi bien que l'égalité des rapports des produits des segments qu'elles interceptent sur les jambes de deux angles à jambes parallèles (théorème appelé plus tard celui de Newton) et même certains théorèmes essentiellement identiques à ceux qui expriment comment on engendre les coniques au moyen de faisceaux projectifs. Dans mon livre, j'ai signalé les analogies que, malgré leur forme géométrique, beaucoup des opérations d'Apollonius présentent avec les procédés algébriques dont on se sert aujourd'hui dans la géométrie analytique; d'autres font pourtant un usage plus direct des résultats géométriques déjà acquis.

En même temps qu'Apollonius profite de l'algèbre géométrique qui existait déja, la nouvelle théorie qu'il développe met à sa disposition de nouveaux moyens algébriques, si toutefois on regarde comme algébriques les solutions dépendant de l'intersection de deux coniques suffisamment déterminées (voir p. 61). Comme moyens de trouver des valeurs numériques, ces solutions laissent beaucoup à désirer; mais pour discuter les problèmes et distinguer des cas où elles sont possibles ou non et trouver ainsi des maxima et des minima, les solutions de cette nature présentent des avantages semblables à ceux de l'algèbre moderne. A ces discussions servaient les déterminations des tangentes, qui permettent de trouver les conditions de con-

tact des deux coniques dont les intersections fournissent la solution d'un problème. Ce sont de telles discussions qu'Apollonius signale, lorsqu'il parle de l'utilité de ses théorèmes pour les diorismes ou délimitations. A côté de cela, il préconise les théorèmes de son livre III comme étant particulièrement utiles pour trouver les solutions de problèmes solides au moyen de lieux solides, dont la nature très générale ressort soit des théorèmes soit des lieux qu'il cite à titre d'exemples. En déterminant, dans le livre IV, le nombre maximum des intersections de deux coniques — y compris l'ensemble de deux branches d'une hyperbole — il fournit un moyen général d'indiquer le nombre total des solutions d'un problème solide.

Cependant, dans les quatre premiers livres de l'œuvre d'Apollonius, destinés à contenir les éléments de la théorie des coniques, c'est-à-dire un exposé systématique de leurs propriétés, on ne trouve pas encore la solution et la discussion directe de problèmes solides concrets ni la détermination des lieux solides dont on y fait usage. Ces solutions et déterminations sont donc réservées pour des recherches particulières. On en trouve une des plus belles dans le livre V, où il s'occupe de la recherche des normales à une conique qui passent par un point donné. Il les détermine (voir le chap. XIII de mon livre cité) au moyen des intersections de la conique avec un autre lieu solide. Malgré son élégance géométrique une telle solution ne possède pourtant pas la régularité d'une solution algébrique moderne, qui fait dépendre le problème d'une seule équation à une seule inconnue, tandis que l'usage de deux lieux solides laisse une plus grande liberté au choix de l'inventeur. Il nous est pourtant conservé une solution d'un autre problème solide que — de même qu'on avait réduit déjà

les problèmes plans à des équations du second degré — on réduit à une équation mixte du troisième degré, qu'ensuite on résout par l'intersection de deux coniques. C'est, en effet, une solution de ce genre qu'on a donné à un problème d'Archimède, demandant de diviser par un plan une sphère en deux segments qui aient un rapport donné. Archimède réduit (Sur la sphère et le cylindre II, 4)[1] le problème à la »section« qui décompose une droite (a) en deux parties (x et $a-x$) telles que

$$\frac{x^2}{b^2} = \frac{c}{a-x}, \qquad (1)$$

où b et c ont des valeurs données. Archimède réserve la solution de ce dernier problème »pour la fin« du même livre mais remarque déja ici qu'en général il »a un diorisme«, qui est d'ailleurs sans importance pour son application au problème particulier considéré. La fin promise du livre était toutefois perdue dès une époque assez reculée; mais Eutocius croit l'avoir retrouvée dans un ancien manuscrit dont il a restitué le contenu, qui est l'une des plus intéressantes recherches conservées de la géométrie grecque[2]. Placé à la fin du livre d'Archimède il aurait pu servir aussi à résoudre la question suivante: trouver un segment de sphère dont le volume et la surface courbe sont donnés, question qu'Archimède ne propose pas directement, mais le dernier théorème de la partie conservée du même livre II en énonce le diorisme qui, suivant la coutume des anciens, devait précéder le problème. Tandis que dans II, 4, l'équation est exprimée par la proportion (1), elle prend plusieurs fois

[1] Voir le chap. XI de mon livre cité et ma »Note sur la résolution géométrique d'une équation du troisième degré par Archimède« (Bibliotheca Mathematica 7 [2] 1903). J'y expose quelques raisons qui me paraissent militer en faveur de l'attribution à Archimède lui-même du manuscrit retrouvé et restitué par Eutocius.

[2] Archimedis Opera, éd. Heiberg III (première édition) p. 154 s.

dans le manuscrit d'Eutocius la forme d'une égalité de deux volumes

$$x^2(a-x) = b^2c, \qquad (2)$$

ou bien on dit qu'il s'agit de partager un segment de droite (a) en deux segments tels que le parallélépipède ayant pour base le carré (x^2) sur l'un, pour hauteur l'autre, ait un volume donné. On remarquera bien l'analogie de cette forme d'une équation du troisième degré avec celle que depuis les Pythagoriciens on donnait à une équation du second degré.

La solution se rattache du reste à la forme (1) de l'équation. Si l'on égale les deux rapports de cette proportion à $\frac{y}{e}$, où e est une quantité arbitrairement donnée, on trouvera le point (x, y) par l'intersection de deux coniques:

$$b^2y = ex^2 \quad \text{et} \quad y(a-x) = ce.$$

Grâce aux propriétés connues des tangentes des deux coniques, on trouve ensuite que le volume $x^2(a-x)$ aura sa valeur maximum si $x = \frac{2a}{3}$, et que le problème aura 2, 1, 0 solutions (positives) suivant que le volume donné est $\lessgtr \frac{4}{27}a^3$.

C'est aux mêmes questions que se rattachent les solutions arabes d'équations du 3e degré au moyen de deux sections coniques. Al Kûhî applique la méthode aux mêmes questions dont nous venons de parler, y compris celle de trouver un segment sphérique dont on connaît le volume et la surface courbe, et par ses discussions qui dépendent de la condition de contact des deux coniques dont il fait usage, il s'élève à la hauteur des auteurs grecs. Plus tard al-Chaijâmî a énuméré les différentes formes des équations du 3e degré et y a ajouté les règles pour leur résolution au moyen de sections coniques. Toutefois, en négligeant la distinction entre les différentes formes d'équations cubiques

que sépare le cas où deux racines deviennent égales de la manière indiquée, distinction identique à celle qu'on a faite plus tard entre les cas dits réductible et irréductible, il ne semble pas avoir remarqué ce qui donne entre les mains des Grecs et d'al Kûhî une véritable valeur théorique aux solutions graphiques dont nous parlons. Peut-être l'énumération de l'auteur éloquent, al Chaijâmî, a-t-elle contribué d'autant plus à exciter le désir d'avoir des solutions d'une autre nature et à soutenir les efforts d'en trouver, jusqu'à ce que cela réussît aux Italiens du commencement du XVI^{e} siècle.

D'ailleurs, les grands progrès des calculs numériques que l'usage du système indien de numération permettait aux Arabes ont contribué à rendre les opérations algébriques plus indépendantes, pratiquement, de la représentation géométrique. Al-Karchî, par exemple, opère avec des radicaux sans faire usage d'une telle représentation. Et les mêmes propriétés caractérisent la forme sous laquelle les opérations algébriques furent transplantées en Italie par Léonard de Pise.

Forelagt paa Mødet d. 24. Januar 1919.
Færdig fra Trykkeriet d. 11. Oktober 1919.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **II**, 5.

TALET

INLEDNING TILL TEORIEN FÖR ANALYTISKA FUNKTIONER

AF

G. MITTAG-LEFFLER

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1920

Efterföljande uppsats är ämnad utgöra de tre första kapitlen af en framställning af Weierstrass' teori för de analytiska funktionerna. Huru nära framställningen än ansluter sig till Weierstrass' tankegång, är den dock icke en reproduktion af någon af hans föreläsningar. Jag har förnämligast eftersträfvat att låta sammanhanget mellan Weierstrass' grundläggande uppfattning och Cantors senare arbeten klart framträda.

Uppfattningen af begreppet helt tal är antagligen icke densamma som hos Weierstrass. Ett litograferadt häfte »Einleitung in die Arithmetik. Erster Teil«, hvilket synes innehålla ett af de mest autentiska referaten af en föreläsning af Weierstrass, inledes med följande ord: »Die Arithmetik hat keine andere Voraussetzung als den Begriff der Zahl. Was aber die Zahl ist, machen wir uns klar, indem wir uns vergegenwärtigen, was wir thun, wenn wir zählen, denn die Zahl ist das Resultat des Zählens«. I en not hänvisas till ett arbete af Zeller, »Logik und Erkenntnistheorie, I, § 13«.

Denna upfattning torde hos Weierstrass alltjämt ha varit den grundläggande. Det förefaller mig emellertid, som om den tankegång, hvilken jag omfattar, bättre träffar talets af all yttre erfarenhet oafhängiga och rätta väsen. Man vinner genom densamma utom annat, att oändlighets-

begreppet samtidigt med talbegreppet framträder omedelbart klart och aprioriskt gifvet. Utgår man däremot vid fastställande af begreppet tal från det ur yttre erfarenhet abstraherade begreppet antal, blir oändligheten oförklarad, och det Weierstrassiska systemet förlorar härmed sitt viktigaste grundlag.

Efter Cantor har mycket debatterats, huruvida icke det grundläggande begreppet helt tal kunde ersättas af ett allmännare, mer omfattande första begrepp, exempelvis mängd. Så intressanta och lockande de försök också äro, hvilka härstämma från en sådan sträfvan[1], synas de mig dock från början vara förfelade. Mängd är icke liksom tal ett aprioriskt gifvet begrepp. Det förutsätter en definition, vid hvilken man nödgas taga till hjälp andra, lika litet aprioriskt gifna begrepp. Detsamma blir fallet vid alla andra försök att ersätta talet med något annat grundbegrepp för tänkandet.

Ett uttryck för denna uppfattning har jag gifvit i de ord

»Talet är tänkandets början och slut
Med tanken föddes talet
Utöfver talet når tanken icke«,

hvilka jag låtit inrista vid ingången till det hem för matematisk forskning, hvilket blifvit uppfördt i Djursholm.[2]

En annan, ehuru oväsentlig, afvikelse från Weierstrass är, att han vid definition och framställning af lagarna för det inkommensurabla talet i stället för hvad jag betecknat med gruppteckned (( )), begagnar sig af det redan i analysen hemmahörande summationstecknet. Jag har valt

[1] Cf. exempelvis E. Borel, »Leçons sur la théorie des fonctions.« 2e éd. Note IV. Paris 1914.

[2] Makarna Mittag-Lefflers matematiska institut. Cf. Acta Mathematica. Bd. 40. 1916. P. III—X.

tecknet (( )) för att den redan skolade matematikern lättare må undvika de felslut, till hvilka ett invandt teckensystem eljest lätt kan ge anledning.

KAPITEL I

Det hela talet.

1. Talet, det hela talet är ett enkelt, aprioriskt begrepp, som icke i ord eller afbild kan definieras, men utgör själfva tänkandets grundform.

Begreppet helt tal erhålles ur en inre åskådning därigenom, att man i föreställningen fasthåller ett gifvet ting, enheten, det första talet och så än en gång samma ting, hvarigenom det andra talet uppkommer, och härefter än en gång samma ting, hvarigenom det tredje talet blifver till, samt härefter fortskrider på samma sätt.

De hela tal, som blifvit bildade medelst samma enhet, sägas tillhöra samma talsystem. De följa hvarandra i en bestämd ordning, så att man om hvarje tal kan angifva, hvilket som är det föregående och hvilket som är det efterföljande. Mot hvarje tal, utom det första, svarar ett närmast föregående annat tal, och mot hvarje tal, inbegripet det första, svarar ett närmast efterföljande annat tal.

Den process, genom hvilken talen bildas, har icke något slut. Man kan alltid fortsätta densamma utöfver den punkt, där man en gång stannat, det finnes icke något tal, som icke har ett efterföljande. I detta förhållande möta vi redan vid tänkandets början oändlighetsbegreppet i dess första form, det matematiska oändlighetsbegreppet, det enda

oändlighetsbegrepp, som för tanken är fattbart. Detta oändlighetsbegrepp är gifvet i och med begreppet helt tal och framstar för tänkandet med lika klarhet som talbegreppet själft.

Begreppen likhet och olikhet beträffande tal, som tillhöra samma talsystem, fastställa vi så, att talen a och b endast då äro lika hvarandra, när de äro samma tal. Detta tecknas $a = b$ eller $b = a$. Talen äro däremot olika hvarandra, om det ena talet icke är detsamma som det andra. Detta tecknas $a \neq b$ eller $b \neq a$. Om a föregår b, eller b efterföljer a, tecknas detta $a < b$ eller $b > a$.

2. Vi ha hittils låtit enheten vara ett visst ting hvilket som hälst. Om man vid detta ting icke fäster någon annan bestämning än att vara det, som vid talets bildande fasthålles i föreställningen, betecknas enheten med ett, och de tal, hvars enhet är ett, benämnas absoluta tal.

Begreppen större och mindre äro icke för absoluta tal aprioriska begrepp, hämtade från någon allmän storhetsteori. De vinnas genom definition på sådant sätt, att af två med hvarandra olika tal är det föregående det mindre och det efterföljande det större.

Teoremet »att det ej finnes något tal, som icke har ett efterföljande«, kommer härigenom att för absoluta tal kunna uttalas så, att »det icke finnes något tal, som är större än alla de öfriga«.

3. En gång i besittning af begreppet helt tal, kommer man omedelbart genom definition till begreppet summa af två tal. »Summan af tvänne tal är det nya tal, som erhålles, om man fortsätter det upprepande af enheten, som ägde rum vid det första talets bildande, lika ofta som det upprepande af enheten, hvilket erfordras för att bilda det andra talet.« Man kan också uttrycka detta så: »Sum-

man af ett gifvet tal och enheten är det tal, hvilket närmast efterföljer det gifna talet. Summen af tvänne tal, af hvilka det sista icke är enheten, är åter det tal, som närmast efterföljer det, som är summan af det första talet och det tal, som närmast föregår det sista.«

Ur vår definition på summa följer omedelbart: »Summan af tvänne tal är alltid samma entydigt gifna tal.«

Denna definition på summa framträder visserligen formelt som en enda definition, men är i verkligheten en obegränsad kedja af definitioner, en definition för hvarje talkombination. Detta är ett förhållande, som vi i det följande öfverallt skola återfinna och som utom annat ger den matematiska analysen medel att under sitt välde lägga städse nya områden af obegränsad omfattning.

4. Sedan begreppet summa en gång blifvit fastställdt, kunna tvänne hufvudteorem härledas, ur hvilka en summas öfriga egenskaper utan vidare härflyta. Det första hufvudteoremet, som innehåller den s. k. associationslagen, lyder:

»Om a, b, c äro trenne gifna tal, så är summan af talet a och det tal $b + c$, hvilket utgör summan af b och c, lika med summan utaf talet $a + b$ — det tal, som är summan af a och b, — samt c, eller

$$a + (b + c) = (a + b) + c.«$$

Låt oss nämligen med e beteckna enheten för de tal, som äro i fråga. Om talet c är enheten e, innehåller associationsteoremet intet annat än ett upprepande af definitionen på summa af två tal. Låt oss nu antaga, att satsen gäller för talen a, b och c. Den gäller då äfven för a, b och $c + e$, det vill säga man har $a + [b + (c + e)] = (a + b) + (c + e)$.

Ty

$$\begin{aligned} a+[b+(c+e)] &= a+[(b+c)+e] && \text{(def. på summa)} \\ &= a+(b+c)]+e && (\ \text{»}\ \ \text{»}\ \ \text{»}\ ) \\ &= [(a+b)+c]+e && \text{(vårt antagande)} \\ &= (a+b)+(c+e) && \text{(def. på summa)} \end{aligned}$$

Nu gäller vår sats, då c är enheten e, den gäller således äfven då c är $e+e$, således också, då c är $e+e+e$ etc. och således för hvarje tal c, huru talen a och b än må ha blifvit valda.

Vi träffa här för första gången ett sätt att sluta, som plägar kallas den fullständiga induktionen eller slutsättet från n till $n+1$. Detta sätt att sluta eller denna bevismetod är följande. Man bevisar, att om en sats gäller för ett tal n, hvilket som hälst, så gäller det äfven för närmast följande tal $n+1$. Nu gäller satsen för ett visst tal $\bar{n}$, i vårt föregående fall talet 2, den gäller då för $\bar{n}+1$ och gäller vidare för härpå närmast följande tal och fortsättningsvis eller med andra ord för alla tal, som följa efter det gifna talet $\bar{n}$.

Detta slutsätt från n till $n+1$ är en af matematikens främsta och oundgängliga bevismetoder.

Vårt andra teorem innehåller den s. k. commutationslagen och lyder: »Om a och b äro tvänne tal inom samma talsystem, så är summan af a och b lika med summan af b och a, eller $a+b=b+a$.«

Om e liksom förut är enheten i talsystemet, så gäller att börja med satsen uppenbarligen, om $a=e$ och samtidigt $b=e$. Vidare är tydligt, att om teoremet gäller för $b=e$ samt ett gifvet a, så gäller det äfven för $b=e$ och $a+e$.

Man har nämligen:

$$\begin{aligned}(a+e)+e &= (e+a)+e \qquad \text{(antagande)}\\ &= e+(a+e) \qquad \text{(def. på summa)}\end{aligned}$$

Då nu teoremet gäller för $b = e$ och $a = e$, gäller det följaktligen äfven för $b = e$ och ett hvilket som hälst tal a.

Om nu åter teoremet gäller för ett a, hvilket som hälst, och ett gifvet b, gäller det äfven för samma a och $b + c$. Man har nämligen:

$$\begin{aligned}a+(b+e) &= (a+b)+e \qquad \text{(def. på summa)}\\ &= (b+a)+e \qquad \text{(antagande)}\\ &= b+(a+e) \qquad \text{(def. på summa)}\\ &= b+(e+a) \qquad \text{(nyss bevisadt)}\\ &= (b+e)+a \qquad \text{(associationslagen)}\end{aligned}$$

Teoremet gäller för ett a hvilket som hälst och för $b = e$. Det gäller således också nu för ett a hvilket som hälst och ett b hvilket som hälst. H. s. b.

5. Genom kombination af associations- och commutationslagarna erhåller man nu utan svårighet, dels:

»Om vissa gifna tal skola summeras, är det likgiltigt, i hvilken ordning summationen företages. Slutresultatet blir alltid ett och samma tal;«

dels äfven:

»Om vissa gifna tal skola summeras, kan man godtyckligt indela talen i grupper; då gruppen innehaller flera tal än ett, bilda summan af talen i hvarje grupp, samt sedan summera alla de så erhållna talen. Man erhåller alltid till slutresultat samma tal.«

6. Sedan vi utredt begreppet summa, blir det lätt fastställa begreppet skillnad mellan tvänne tal. Man härleder nämligen lätt satsen: »Om tvänne olika tal äro gifna, så finnes alltid ett entydigt gifvet tredje tal i samma talsystem (= skillnaden mellan det efterföljande och det

föregående af de gifna talen) sådant, att summan af det föregående af de gifna talen och det tredje talet är lika med det efterföljande af de båda talen« eller algebraiskt uttryckt, »likheten

$$a + x = b,$$

där a och b äro de båda gifna talen och a föregår b, har alltid till lösning ett entydigt gifvet tal x (= skillnaden mellan b och a)«.

Vi anteckna följande satser, hvilka härur omedelbart härflyta.

»Om a, b, c äro tre gifna tal i samma talsystem och $a + b = c$, så är $a < c$ och $b < c$.«

»Om a, b, c, d äro tal i samma talsystem och $a = b$, $c = d$, så är $a + c = b + d$.«

»Om $a > b$ och $c > d$, så är $a + c > b + d$.«

»Om $a = b$ och $c > d$, så är $a + c > b + d$.«

samt omvändt:

»Om $a = b$ och $a + c > b + d$, så är $c > d$.«

»Om $a = b$ och $a + c = b + d$, så är $c = d$.«

»Om $a > b$ och $a + c = b + d$, så är $c < d$.«

7. Vi öfverga nu till en utredning af begreppet produkt af tal. Vi definiera på följande sätt produkten af ett tal och ett absolut tal: »Med produkten af ett tal a och ett absolut tal n förstås, då n är större än ett, summan af n tal a och, då n är lika med ett, talet a självt. Denna produkt tecknas $a \cdot n$.«

Om således e är enheten i talsystemet och n är det antal ganger enheten upprepats vid talets bildande, kommer talet att tecknas $e \cdot n$. Är enheten ett, blir således talet $1 \cdot n$ och, emedan detsamma äfven tecknats n, erhålles likheten

$$1 \cdot n = n.$$

Man inser nu utan vidare förklaring betydelsen af uttrycket:

$$(((a \cdot m_1) \cdot m_2) \cdot m_3) \ldots m_n,$$

där m_1, m_2, $m_3 \ldots m_n$ äro absoluta tal. Det genom detsamma framställda talet skrifves enklare

$$a \cdot m_1 \cdot m_2 \cdot m_3 \ldots m_n.$$

Man inser likaså betydelsen af uttrycket

$$a \cdot m_1 \cdot m_2 \cdot m_3 \ldots m_n \cdot (l_1 \cdot l_2 \ldots l_p),$$

där l_1, l_2, $\ldots$ l_p, liksom förut m_1, $m_2 \ldots m_n$ äro absoluta tal.

Man ser vidare, att om a_1, $a_2 \ldots a_n$ äro tal inom samma talsystem och m_1, $m_2 \ldots m_p$ äro absoluta tal, så är:

$$\begin{aligned}(a_1 + a_2 + \ldots + a_n)\ (m_1 + m_2 + \ldots + m_p) = \\ a_1 \cdot m_1 + a_2 m_1 + \ldots + a_n \cdot m_1 + \\ a_1 \cdot m_2 + a_2 \cdot m_2 + \ldots + a_n \cdot m_2 + \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_1 \cdot m_p + a_2 \cdot m_p + \ldots + a_n \cdot m_p.\end{aligned}$$

8. Vidare ser man, att associations- och commutationslagarna äfven gälla för produkten af tal. De komma härvid att lyda:

»Låt a vara ett gifvet tal, och m och n tvänne gifna absoluta tal. Produkten af talet $a \cdot m$ samt talet n är lika med produkten af talet a och produkten af talen m och n, eller

$$(a \cdot m) \cdot n = a \cdot (m \cdot n).«$$

Detta är associationslagen och ur densamma härflyter:

$$»a \cdot m_1 \cdot m_2 \ldots m_p = a \cdot (m_1 \cdot m_2 \ldots m_p),$$

om med m_1, $m_2 \ldots m_p$ förstås absoluta tal.«

Commutationslagen blir åter:

»Låt oss med m och n förstå absoluta tal. Produkten af m och n är lika med produkten af n och m, eller

$$m \cdot n = n \cdot m.«$$

Ur commutationslagen följer den allmännare satsen: »Om vissa bestämdt angifna absoluta tal skola multipliceras med hvarandra, är det likgiltigt i hvilken ordning multiplikationen företages. Slutprodukten blir alltid samma tal. Man erhaller också samma tal till slutprodukt, om man huru som hälst indelar de gifna talen i grupper, bildar produkten af talen i hvarje grupp, som innehaller mer än ett tal, samt härefter multiplicerar med hvarandra dessa produkter och, om dylika finnas, de tal, som äro ensamma i sin grupp.«

9. Vi anteckna äfven följande satser, hvilka motsvara de satser, som vi i § 6 sammanställt för summor af tal. De erhållas utan svårighet ur det föregående.

»Låt m, n, l vara tränne absoluta tal och antag, att $m \cdot n = l$. Om hvardera af de tre talen är större än ett, är $m < l$ och $n < l$. Om ettdera af talen m och n är ett, är det andra af dem lika med l, och om talet l är ett, äro båda talen m och n lika med ett.«

Låt oss nu med a och b förstå två tal i samma talsystem samt med m och n tvänne absoluta tal. Då gälla satserna:

»Om $a = b$ och $m = n$, så är $a \cdot m = b \cdot n$.«

»Om $a > b$ och $m > n$, så är $a \cdot m > b \cdot n$.«

»Om $a = b$ och $m > n$, så är $a \cdot m > b \cdot n$.«

samt omvändt satserna:

»Om $a = b$ och $a \cdot m > b \cdot n$, så är $m > n$.»

»Om $a = b$ och $a \cdot m = b \cdot n$, så är $m = n$.«

»Om $a > b$ och $a \cdot m = b \cdot n$, så är $m < n$.«

10. Uttrycken, att ett tal är en mångfald af ett annat tal eller en divisor till ett tal, bestämmas på följande sätt.

»Om med a förstås ett gifvet tal och med n ett gifvet absolut tal samt med b betecknas talet $a \cdot n$ ($b = a \cdot n$), så säges talet b vara den n'te mångfalden af talet a, och talet a sägas vara en divisor till b.«

Med begagnande af denna terminologi kan man uttala de båda satserna:

»Enheten är en gemensam divisor för alla tal i talsystemet.«

»Hvarje tal har sig självt till divisor.«

samt omedelbart härleda satsen:

»Om $a = b + c$ och d är divisor till två af talen a, b, c, så är d divisor äfven till det tredje talet.«

11. Låt oss nu antaga, att a och b äro tvänne med hvarandra olika tal, af hvilka det ena icke är divisor till det andra. I sjunde boken af Euklides Elementa läres, huru man skall finna en divisor till såväl a som b, hvilken själf som divisor innehåller hvarje tal, som är divisor till både a och b. Från detta Euklides förfarande stamma en hel rad af den högre aritmetikens och algebrans viktigaste satser. Detta förfarande är i hufvudsak följande.

Låt oss med a förstå det föregående och med b det efterföljande talet. Om man bildar de på hvarandra följande mångfalderna $a \cdot 1$, $a \cdot 2$, $a \cdot 3$ ... träffar man nödvändigt en första mångfald af a, som är ett tal, hvilket följer efter b och som således närmast föregås af en mångfald af a, låt vara $a \cdot \beta$, hvilken föregår b. Det finnes då nödvändigt (§ 6) ett och endast ett tal c, hvilket är sådant att

$$b = a \cdot \beta + c.$$

Detta tal c föregår talet a. Om man nu å c och a upprepar samma förfarande som förut å a och b, erhåller man:

$$a = c \cdot \gamma + d,$$

där γ är ett absolut tal i samma system som a, b, c, hvilket föregår c.

Upprepar man ytterligare detta förfarande, erhåller man en följd af likheter:

$$\begin{aligned} b &= a \cdot \beta + c \\ a &= c \cdot \gamma + d \\ &\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots \\ s &= t \cdot \tau + u, \end{aligned}$$

där a föregår b, c föregår a, d föregår c och u föregår t. Man framkommer slutligen till en sista likhet

$$t = u \cdot \eta,$$

hvilken uttrycker, att det sista talet u är en divisor till det näst sista talet t.

Emedan (§ 10) hvarje divisor till a och b äfven är en divisor till c och härmed till $d \ldots s$, t, u samt dessutom u är en divisor till t, $s \ldots d$, c, b, a, ha vi uti u erhållit ett tal, hvilket, självt divisor till såväl a som b, som divisor innehåller hvarje annan gemensam divisor till a och b. Det kan ej heller finnas något annat tal med denna egenskap, ty ett tal, som föregår u, kan icke ha u till divisor, och u kan icke heller som divisor ha ett tal, hvilket efterföljer u.

Vi kunna således uttala följande sats, som är andra propositionen i sjunde boken af Euklides Elementa.

»Om a och b äro tvänne med hvarandra olika tal i samma talsystem, så finnes alltid en gemensam divisor till båda talen, hvilken själf som divisor innehåller hvarje divisor till de båda talen. Det finnes icke mer än ett tal,

som har denna egenskap, och detsamma erhålles genom Euklides förfarande.«

12. Om a och b äro absoluta tal, är u den största gemensamma divisorn till a och b, och Euklides' förfarande har således gifvit oss medel att finna denna största gemensamma divisor.

»Om denna största gemensamma divisor är ett, sägas de båda talen vara relativa primtal till hvarandra.«

Man är nu i stånd att utan vidare härleda följande båda satser.

»Om a och b äro relativa primtal till hvarandra, så är det minsta tal, hvilket är en mångfald såväl af talet a som af talet b talet $a \cdot b$.«

»Om d är den största gemensamma divisorn till tvänne tal, $a = d \cdot a'$ och $b = d \cdot b'$, så är det minsta tal, hvilket är en mångfald såväl af talet a som af talet b, talet $d \cdot a' \cdot b'$.«

Detta tal $d \cdot a' \cdot b'$ benämnes den minsta gemensamma dividenden till talen a och b.

13. Om vi nu fortsatte studiet af de hela talens addition och multiplikation, skulle vi komma till talteorien, en teori som har sin källa redan i Euklides förfarande och som sedan genom Diophantus, genom Fermat och Euler, Gauss, Lejeune Dirichlet och Riemann samt deras efterföljare, nu senast genom Minkowski, vunnit en sådan betydenhet, att ett yttrande, hvilket tillskrifves Gauss, att talteorien vore »die Königin der Mathematik« på samma sätt som matematiken vore »die Königin der Wissenschaften«, väl kan vara värdt att begrunda.

Man hade då att utgå från primtalen, »där med primtal förstås ett absolut tal, som icke har andra divisorer än talet självt samt ett«, och att härefter börja med att bevisa den grundläggande satsen, att »hvarje tal är en pro-

dukt af primtal« samt att »när tva primtalsprodukter framställa samma tal, hvarje primtalsfaktor i den ena produkten förekommer lika ofta i den andra.«

KAPITEL II

Det brutna talet.

1. Om med a förstås ett gifvet tal, kan man, som vi sett, alltid bilda ett nytt tal b, hvilket utgör en viss bestämd mangfald af a. Om åter b är ett gifvet tal, finnes icke nödvändigt ett motsvarigt tal sadant, att en viss mångfald af detsamma är lika med b. Likheten $x \cdot n = b$, där x är ett tal i samma talsystem som b och n är ett absolut tal, kan icke uppfyllas för hvilka värden som hälst på n och b.

För att kunna upprätthålla en sådan sats, att denna likhet vid hvilka tal som hälst n och b alltid har en lösning, blir det nödvändigt ålägga enheten den bestämning, att densamma är delbar, och vidga vårt talsystem därhän, att detsamma kommer att innehålla äfven tal, som äro bildade af delar af enheten.

Vi ha hittils icke pålagt enheten e någon bestämning utöfver den att vara det ting, som vid talets bildande fasthalles i föreställningen. Vi ålägga nu enheten härutöfver egenskapen, »att, huru vi än fixera ett absolut tal n, städse vara den n'te mångfalden af en annan i och med n entydigt gifven enhet. Denna sista enhet benämnes den n:te brakdelen af den ursprungliga enheten och tecknas, då med e förstås den ursprungliga enheten, $\frac{e}{n}$.«

Vi uttrycka detta äfven så, »att vi om enheten e förutsätta, att densamma städse och vid hvarje val af det ab-

soluta talet n är delbar och endast på ett sätt delbar i n med hvarandra equivalenta delar.«

Vid hvarje likhet eller olikhet, där den ursprungliga enheten e uppträder, skall man således kunna ersätta densamma med en n:te mångfald (det absoluta talet må vara hvilket som hälst) af enhetens n:te bråkdel $\left(e = \frac{e}{n} \cdot n\right)$ och öfver allt, där den n:te mångfalden af den n:te bråkdelen af den ursprungliga enheten förekommer, skall denna n:te mångfald kunna ersättas af den ursprungliga enheten $\left(\frac{e}{n} \cdot n = e\right)$.

2. Det är lätt att se, att hvarje bråkdel af enheten åter själf är delbar i ett huru stort antal delar som hälst. För att inse detta erfordras endast att ådagalägga, att huru man än må välja de absoluta talen m och n, finnes alltid en bråkdel af enheten e, hvars m:te mångfald är lika med $\frac{e}{n}$. Om bråkdelen $\frac{e}{n \cdot m}$ gäller $\frac{e}{n \cdot m} \cdot n \cdot m = e$ och om bråkdelen $\frac{e}{n}$, att $e = \frac{e}{n} \cdot n$. Emedan nu $\frac{e}{n \cdot m} \cdot n \cdot m = \frac{e}{n \cdot m} \cdot m \cdot n$ (Kap. I, § 10), så äro såväl den n:te mångfalden af $\frac{e}{n \cdot m} \cdot m$ som af $\frac{e}{n}$ lika med e. Då vi nu förutsatt, att den n:te bråkdelen af e eller den enhet, hvars n:te mångfald är e, är entydigt gifven, så är således $\frac{e}{n \cdot m} \cdot m = \frac{e}{n}$ och $\frac{e}{n \cdot m}$ är i följd häraf den m:te bråkdelen af $\frac{e}{n}$

$$\left(\frac{\frac{e}{n}}{m} = \frac{e}{n \cdot m}\right).$$

Då nu $\frac{e}{n \cdot m} = \frac{e}{m \cdot n}$ erhåller man således äfven satsen

$$\text{»}\frac{\frac{e}{n}}{m} = \frac{\frac{e}{m}}{n}\text{«}.$$

»Vi beteckna hädanefter vid det talsystem, hvars enhet är e denna enhet som systemets grundenhet och anse talsystemet omfatta såväl alla tal, bildade af enheten e som äfven alla tal $\frac{e}{n}\cdot m$, bildade af en bråkdel af enheten e, hvilka icke samtidigt äro en multipel af denna enhet. De förra talen betecknas då som hela tal och de senare som brutna tal.«

Vid enheten ett, som förut icke hade någon annan egenskap än att vara det, som vid talets bildande fasthölls i föreställningen, fästes således numera ytterligare den egenskap, att »densamma är delbar« och »vårt absoluta talsystem kommer därföre att hädanefter omfatta såväl hela som brutna tal.«

Man har vidare

$$»1 = \frac{1}{n}\cdot n \text{ och } \frac{1}{n}\cdot n = 1.«$$

Vi anteckna ytterligare

$$»\frac{e}{1} = e \text{ och } \frac{1}{1} = 1.«$$

»Talet $\frac{e}{n}\cdot m$ tecknas äfven $\frac{e\cdot m}{n}$ eller, då $e\cdot m = b$, likaväl $\frac{b}{n}$.«

»Detta tal $\frac{e}{n}\cdot m$ är den n:te bråkdelen af talet b«, ty

$$\frac{b}{n}\cdot n = \frac{e\cdot m}{n}\cdot n = \frac{e}{n}\cdot m\cdot n = e\cdot m.$$

»Likheten $x\cdot n = b = e\cdot m$, som utgjorde vår utgångspunkt är således alltid lösbar och lösningen är

$$x = \frac{b}{n} = \frac{e\cdot m}{n} = \frac{e}{n}\cdot m.«$$

»Talet $\frac{b}{n} = \frac{e\cdot m}{n}$ benämnes bråk och säges härvid $b = e\cdot m$ vara bråkets täljare samt det absoluta hela talet n dess

nämnare.« »När man talar om bråk, utan att angifva, att man befinner sig inom ett talsystem med grundenheten e, betyder detta, att $e = 1$ eller att man har ett tal inom det absoluta talsystemet.«

3. Emedan $a = b$ icke längre betyder, att a är identiskt samma tal som b, utan tvärtom a kan uppträda under formen $\frac{e \cdot m \cdot p}{n \cdot p}$ och b under formen $\frac{e \cdot m \cdot q}{n \cdot q}$, hvarvid $p \neq q$, blir det, för att kunna med hvarandra jämföra tvänne tal, nödvändigt att först bringa dem under enhetlig form. Detta sker så, att man först reducerar hvardera talet till sadan form, att de absoluta hela talen i täljare och nämnare blifva relativa primtal. Om sedan detta skett, det ena talet befinnas vara bildadt af μ:te delen och det andra af ν:te delen af enheten, uppsöker man den minsta gemensamma dividenden γ till μ och ν och reducerar sedan hvardera talet till att blifva en mångfald af $\frac{e}{\gamma}$. Härigenom blir det möjligt att afgöra om talen äro lika eller olika hvarandra, hvilket tal, som bör anses vara det föregående och hvilket det efterföljande, samt i det fall, att $e = 1$, hvilket tal, som vid olikhet mellan tvänne tal är det större och hvilket som är det mindre.

Vi erhålla härigenom:

»Om a, b, c tillhöra samma talsystem och $a = b$ samt $b = c$, så är $a = c$.

Om $a > b$ samt $b > c$, så är $a > c$.

Om $a > b$ och $b = c$, så är $a > c$.

Om $a = b$ och $b > c$, så är $a > c$.

4. Närmast kommer nu att definiera summa af tvänne tal i vårt utvidgade talsystem. Uppenbarligen bör denna definition komma att lyda:

»Med summan af a och b förstas det tal, som erhålles

om båda talen uttryckas som mångfalder af samma bråkdel af talsystemets enhet, samt härefter summeras som tal, hvilka ha denna bråkdel till enhet.«

Man ser lätt, att »den på detta sätt definierade summan $a + b$ blir entydigt bestämd eller oberoende af, i hvilken bråkdel af enheten, man må ha uppmätt de båda talen a och b.«

Man ser vidare, att associations- och commutationslagarna (Kap. I, § 4) fortfarande gälla:

$$\text{»}a + (b + c) = (a + b) + c$$
$$a + b = b + a,\text{«}$$

och att likaså fortfarande våra satser Kap. I, § 5 och § 6 kunna oförändrade upprätthållas.

5. Vi ha förut funnit (Kap. I, § 7), att om $a_1, a_2 \ldots a_n$ äro tal inom samma talsystem och $\alpha_1, \alpha_2, \ldots \alpha_p$ äro absoluta hela tal, så är:

$$\begin{aligned}(a_1 + a_2 + \ldots + a_n)(\alpha_1 + \alpha_2 + \ldots + \alpha_p) = \\ a_1 \cdot \alpha_1 + a_2 \cdot \alpha_1 + \ldots + a_n \cdot \alpha_1 + \\ a_1 \cdot \alpha_2 + a_2 \cdot \alpha_2 + \ldots + a_n \cdot \alpha_2 + \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_1 \cdot \alpha_p + a_2 \cdot \alpha_p + \ldots + a_n \cdot \alpha_p\end{aligned}$$

För att kunna upprätthålla denna sats äfven för det fall, att talen a och α äro bråk, blir det nödvändigt att definiera, hvad som bör förstås med $\frac{e \cdot m}{n} \cdot \frac{p}{q}$, där m, n, p och q äro absoluta tal. Vi ha:

$$\frac{e \cdot m}{n} \cdot \frac{p}{q} = \left(\frac{e}{n} + \ldots + \frac{e}{n}\right)\left(\frac{1}{q} + \ldots + \frac{1}{q}\right),$$

hvarest $\frac{e}{n}$ förekommer m gånger inom första klammern samt $\frac{1}{q}$ förekommer p gånger inom andra klammern.

Skall vår sats kunna upprätthållas, bör således:

$$\left(\frac{e}{n}+\ldots+\frac{e}{n}\right)\left(\frac{1}{q}+\ldots+\frac{1}{q}\right)=\frac{e}{n}\cdot\frac{1}{q}+\ldots+\frac{e}{n}\cdot\frac{1}{q}+$$
$$\frac{e}{n}\cdot\frac{1}{q}+\ldots+\frac{e}{n}\cdot\frac{1}{q}+$$
$$\ldots\ldots\ldots\ldots$$
$$\frac{e}{n}\cdot\frac{1}{q}+\ldots+\frac{e}{n}\cdot\frac{1}{q}$$

där $\frac{e}{n}\cdot\frac{1}{q}$ i hvarje horizontal rad förekommer m gånger och i hvarje vertikal rad p gånger samt $\frac{e}{n}\cdot\frac{1}{q}$ bör vara ett tal inom det system, hvars grundenhet är e och som det återstår att definiera.

Om $\frac{e}{n}\cdot\frac{1}{q}$ tillhör detta talsystem, är

$$\left(\frac{e}{n}\cdot\frac{1}{q}\right)\cdot q\cdot n=\frac{e}{n}\cdot\frac{1}{q}+\ldots+\frac{e}{n}\cdot\frac{1}{q}+$$
$$\frac{e}{n}\cdot\frac{1}{q}+\ldots+\frac{e}{n}\cdot\frac{1}{q}+$$
$$\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots$$
$$\frac{e}{n}\cdot\frac{1}{q}+\ldots+\frac{e}{n}\cdot\frac{1}{q},$$

hvarest $\frac{e}{n}\cdot\frac{1}{q}$ i hvarje horizontal rad förekommer q gånger och horizontalradernas antal är n.

Nu är:

$$\frac{e}{n}\cdot\frac{p}{q}=\frac{e}{n}\left(\frac{1}{q}+\ldots+\frac{1}{q}\right)$$

där $\frac{1}{q}$ inom klammern förekommer p gånger, och således äfven:

$$\frac{e}{n}\cdot\frac{q}{q}=\frac{e}{n}\left(\frac{1}{q}+\ldots+\frac{1}{q}\right),$$

där $\frac{1}{q}$ förekommer inom klammern q gånger. Således är:

$$\left(\frac{e}{n}\cdot\frac{1}{q}\right)\cdot q\cdot n = \frac{e}{n}\cdot\frac{q}{q}+\ldots+\frac{e}{n}\cdot\frac{q}{q} = \frac{e}{n}\cdot 1+\ldots+\frac{e}{n}\cdot 1,$$

där hvarje term på högra sidan om likhetstecknet förekommer n gånger.

Vi ha således blifvit ledda därtill, att för att vår sats Kap. I, § 9 skall kunna upprätthallas, utan att man lämnar det talsystem, hvars grundenhet är e, erfordras definitionen:

$$\text{»}\frac{e}{n}\cdot\frac{1}{q} = \frac{e}{q\cdot n}\text{«}$$

eller, då $e = 1$,

$$\text{»}\frac{1}{n}\cdot\frac{1}{q} = \frac{1}{n\cdot q}\text{«}.$$

Med denna definition till utgångspunkt erhaller man satsen i fråga och kan således också skrifva:

$$\text{»}\frac{e\cdot m}{n}\cdot\frac{p}{q} = \frac{e}{n}\cdot\frac{1}{q}\cdot m\cdot p = \frac{e}{nq}\cdot m\cdot p = \frac{e\cdot m\cdot p}{n\cdot q}\text{«}.$$

Man inser äfven utan vidare, att satserna Kap. I, § 8 och 9 kunna oförändrade upprätthållas.

6. Vi ha sett, att teorien för de brutna talen har sin källa i sträfvan att upprätthålla lösbarheten af likheten

$$x\cdot\alpha = a = e\cdot m$$

äfven i det fall, då det absoluta talet α icke är en divisor till det absoluta talet m.

Denna sträfvan, hvars yttre anledning är våra sinnens uppfattning om vissa yttre tings delbarhet, går tillbaka till mänskligheten tidigaste barndom, men har relativt sent blifvit uttryckt under den abstrakta form, i hvilken vi klädt densamma. Denna form är emellertid det begreppsliga och adeqvata uttrycket för hvad man mer eller mindre medvetet eftersträfvat.

En af matematikens, den matematiska analysens, främsta uppgifter är äfven att bringa till full klarhet den mer och mer omfattande kunskap om tingen och deras förhållanden, som människorna småningom förvärfva. Det har alltid varit så, huru långt vi än ga tillbaka i tänkandets historia, det är så äfven i våra dagar, när det gäller att begreppsligt fatta och ordna det brokiga och omfattande nya kunskapsmaterial, som naturvetenskaperna alltjämt insamla, och det skall också alltid så förblifva, så länge vår sinnesvärld och vårt tänkande stå i förhållande till hvarandra.

7. Låt oss med a och b $(a < b)$ förstå tvänne tal i vårt utvidgade talsystem (§ 2), hvilka må vara hvarandra huru närbelägna som hälst. Huru stort man än må välja det absoluta talet m, finnas alltid m med hvarandra olika tal, hvilka efterfölja a, men föregå b. Vid första påseende förefaller det därföre, som hade oändlighetsbegreppet (Kap. I, § 1) genom införande af de brutna talen blifvit i mycket väsentlig mån utvidgadt. Att så emellertid icke är förhållandet, inses på följande sätt.

Samtliga absoluta tal kunna uttryckas under formen $\frac{\lambda}{\mu}$. Låt oss indela dessa tal i grupper, så att till första gruppen höra alla tal, för hvilka $\lambda + \mu = 2$ (således endast talet 1), till andra gruppen alla tal, för hvilka $\lambda + \mu = 3$ (således 2, 1/2) till tredje gruppen alla tal, för hvilka $\lambda + \mu = 4$ och hvilka icke förekomma i en af de föregående grupperna (således 3, 1/3) etc., till m:te gruppen alla tal, för hvilka $\lambda + \mu = m + 1$ och hvilka icke förekomma i någon föregående grupp etc. Man inser, att man genom att fortsätta på detta sätt erhåller hela talvärlden. I första gruppen finns endast ett tal, i andra två tal, i tredje också två tal i m:te p tal etc. Låt oss nu till talen 1, 2, 3, ... n ... tillordna, till talet 1 talet 1, till de två

talen 2, 3 ett af hvardera af de båda talen i andra gruppen, till talen 4, 5 ett af hvardera af de båda talen i tredje gruppen samt fortskrida på samma sätt, så att de p talen i m:te gruppen hvart och ett af dem tillordnas ett af talen $p_1 + p_2 + \ldots + p_{m-1} + 1$, $p_1 + p_2 + \ldots + p_{m-1} + 2$, $\ldots$ $p_1 + p_2 + \ldots + p_{m-1} + p_m$. Man erhåller på detta sätt hela vår talvärld entydigt tillordnad talserien $1, 2, 3 \ldots n \ldots$

Vi ha vid detta sätt att gå tillväga för första gången användt tankens förmåga att på en gång i medvetandet fasthalla tvänne föremål, hvilka äro till hvarandra tillordnade.

Utan denna tankens förmåga funnes icke någon matematisk vetenskap.

Vi ha på detta sätt erhållit: **Oändlighetsbegreppet har icke blifvit vidgadt genom införande af de brutna talen.**

KAPITEL III

Grupper af ett obegränsadt antal utaf tal. Det allmänna, absoluta talet.

1. Vi ha sett, huru samtliga hela och brutna tal kunna entydigt tillordnas den obegränsade följden $\mu = 1, 2, 3 \ldots$ af hela tal på ett sådant sätt, att mot hvarje helt tal μ svarar ett och endast ett helt eller brutet tal a_μ samt, vice versa, mot hvarje helt eller brutet tal a_μ ett och endast ett helt tal μ.

Låt oss nu från talen a_μ tänka oss utbrutna vissa gifna tal. Sammanfattningen af dessa tal benämna vi en **grupp** och teckna densamma (a). Vi benämna vidare hvarje enstaka

tal a, som tillhör gruppen, ett element i densamma. Det kan nu inträffa, att huru stort antal element jag frånskiljer gruppen (a), andra element kvartstå. I så fall säges gruppen vara sammanfattningen af ett obegränsadt antal element.

En sådan grupp (a), som utgör sammanfattningen af ett obegränsadt antal element, vore en grupp, hvars element äro samtliga hela och brutna tal. En annan vore sammanfattningen af alla brutna tal mellan två på hvarandra följande hela tal. Ännu en annan vore sammanfattningen af alla tal $\frac{1}{2^\nu}$, där ν antager alla heltalsvärden eller af $\frac{1}{p^\nu}$, där p betyder samtliga primtal och ν alla hela tal.

Vi förenkla att börja med vår undersökning genom förutsättningen, att alla tal i gruppen äro absoluta tal. Senare skola vi behandla allmännare grupper, där detta ej längre är fallet.

Vi börja med att fastställa som grundläggande förutsättning för de grupper, hvilka af oss i det följande skola behandlas: Det finnes alltid vid hvarje grupp ett tal, som är större än summan af ett huru stort antal som hälst af element i gruppen.

Vi teckna en sådan grupp $((a))$. Ingen sådan i förväg bestämd tillordning af elementen a till hvart och ett af talen $\nu = 1, 2, 3 \ldots$, som exempelvis uppkommer, om man uppfattar talen a som tal a_{μ_ν} enligt någon bestämd lag utbrutna ur talföljden a_α, förutsättes härvid som gifven. Ingendera af de båda grupperna, sammanfattningen af alla hela och brutna tal eller af alla brutna tal mellan två hela tal, uppfyller var grundläggande förutsättning. Man kan nämligen alltid bilda en summa af element i en sådan grupp, som är större än hvilket som hälst uppgifvet tal. Ingendera af dessa båda grupper är således en grupp $((a))$.

Detta är däremot fallet med gruppen $\frac{1}{2^\nu}$; $\nu = 1, 2, 3 \ldots$ Man har nämligen, n må vara ett helt tal huru stort som hälst:

$$\sum_{\nu=1}^{n} \frac{1}{2^\nu} = \frac{1 - \frac{1}{2^n}}{1 - \frac{1}{2}} \cdot \frac{1}{2} = 1 - \frac{1}{2^n} < 1.^1$$

Det må ej vara uteslutet, att vissa element i gruppen $((a))$ äro lika med ett och samma tal. Detta kan dock i följd af vår grundläggande förutsättning endast inträffa för ett begränsadt antal element och för hvarje element endast ett begränsadt antal gånger.

Med den grundläggande förutsättning, hvilken vi beträffande grupperna $((a))$ infört, skall det visa sig, att vi i teorien för dessa grupper stå på lika fast mark, som vid studiet af det hela och det brutna talet. Det skall visa sig, att sådana grupper kunna behandlas som tal samt lyda samma räknelagar som de hela och brutna talen. Det skall vidare komma att framgå, att desammas införande innebär ett väsentligt vidgande af talbegreppet, hvarigenom erhålles såväl den begreppsliga tolkningen af ett problem, som under årtusenden gäckat tidernas största tänkare, eller att kvadratens sida och diagonal samt cirkelns periferi och diameter icke kunna mätas med gemensamt mått, som äfven att vägen samtidigt öppnas utöfver det ändligas matematik till infinitesimalkalkylen med alla dess afgreningar, de må benämnas högre matematisk analys, differential- och integralkalkyl, allmän funktionsteori eller teori för analytiska funktioner eller annat.

[1] För enkelhets skull anticipera vi, då härå icke ligger någon särskild vikt, kännedom om betydelsen af tecknet —.

2. Det första, hvilket beträffande våra grupper behöfver fastställas, är begreppen likhet och olikhet samt större och mindre i afseende på tvänne grupper $((a))$ och $((b))$.

Likhet mellan tvänne tal ägde endast rum, när de båda talen genom ett ändligt antal operationer kunde öfverföras uti samma tal. Ett sådant likhetsbegrepp kan icke längre beträffande våra grupper upprätthållas. Begreppet likhet och härmed äfven begreppen större och mindre måste ges en vidare bestämning.

Vi begagna oss af en terminologi af Weierstrass. Ett tal säges innehållas i gruppen $((a))$, om man från gruppen kan frånskilja ett antal element, hvilkas summa är större än talet i fråga. Vi kunna med denna terminologi på följande sätt uttala vår grundläggande förutsättning: En grupp, hvilken icke innehåller samtliga hela tal, är en grupp $((a))$.

Begreppen likhet, större och mindre definiera vi nu på följande sätt: »Gruppen $((a))$ är lika med gruppen $((b))$, $((a)) = ((b))$, $((b)) = ((a))$, om hvarje tal, som innehålles i $((a))$, äfven innehålles i $((b))$ och vice versa.« »Däremot är den ena gruppen $((a))$ mindre än den andra $((b))$, som är större, $((a)) < ((b))$, $((b)) > ((a))$, om det finnes ett tal, hvilket innehålles i $((b))$ utan att innehållas i $((a))$.« Att gruppen $((a))$ är olika gruppen $((b))$, $((a)) \neq ((b))$, betyder då, att likhet mellan de båda grupperna icke äger rum, att de ej äro lika hvarandra.«

Ett exempel på likhet mellan tvänne olika sammansatta grupper erhålles genom jämförelse mellan gruppen $\left(\left(\frac{1}{2^\nu}\right)\right)$; $\nu = 1, 2, 3 \ldots$ samt den grupp, som består af det enda talet 1. Man har, som vi sett, n må väljas huru stort som hälst,

$$\sum_{\nu=1}^{n} \frac{1}{2^\nu} = 1 - \frac{1}{2^n}.$$

Häraf följer, att gruppen $\left(\left(\frac{1}{2^\nu}\right)\right)$ alltid och endast innehaller hvarje egentligt bråk. Då det samma gäller om talet 1, är således

$$\left(\left(\frac{1}{2^\nu}\right)\right) = 1.$$

3. Ur definitionen på *likhet* mellan tvänne grupper följer omedelbart:

»Om gruppen endast innehåller ett begränsadt eller ändligt antal element, så är densamma lika med summan af dessa element«

samt vidare:

»Om $((a)) = ((b))$ och $((b)) = ((c))$, så är $((a)) = ((c))$«.

»Om $((a)) > ((b))$ och $((b)) \geq ((c))$, så är $((a)) > ((c))$«.

»Om $((a)) \geq ((b))$ och $((b)) > ((c))$, så är $((a)) > ((c))$».

4. En enkel öfverläggning visar ytterligare:

»Om δ är ett tal, hvilket innehålles i gruppen $((a))$, så finnes alltid ett motsvarigt helt tal n sådant, att $n \cdot \delta < ((a)) < (n+1)\delta$.« Det finnes nämligen alltid ett så stort helt tal m, att $m \cdot \delta$ är större än hvarje uppgifvet tal. Ökar man tillräckligt m, träffar man därför slutligen på ett tal N ($N > m$) så stort, att $N \cdot \delta$ icke innehålles i $((a))$. Minska nu N med en enhet åt gangen. Man träffar då nödvändigt på ett sista helt tal $n+1$ sådant, att $(n+1)\cdot\delta$ icke innehålles i $((a))$, medan detta däremot är fallet med $n \cdot \delta$. Q. e. d.

Häraf följer åter:

»Från gruppen $((a))$ kan, då densamma innehåller ett obegränsadt antal element, alltid franskiljas så många (= ett så stort antal) element, att den grupp, som här-

efter återstår, är mindre än hvilket uppgifvet tal som hälst.«

Låt nämligen δ vara ett godtyckligt litet tal, hvilket innehålles i $((a))$. Vi ha bevisat tillvaron af ett helt tal n sådant, att

$$n \cdot \delta < ((a)) \leq (n+1) \cdot \delta.$$

Således kan man alltid finna ett så stort helt tal m, att

$$n\delta < \sum_{\nu=1}^{m} a_\nu$$

där a_1, a_2, a_3, $\ldots$ a_m må betyda olika element i gruppen $((a))$, och således äfven ett tal $\delta_1 > \delta$ sådant, att

$$n\delta < n\delta_1 < \sum_{\nu=1}^{m} a_\nu.$$

Häraf följer åter

$$n\delta_1 + \sum_{\nu=m+1}^{m+p} a_\nu < \sum_{\nu=1}^{m+p} a_\nu$$

där a_{m+1}, a_{m+2}, $\ldots$ a_{m+p} äro nya element ur gruppen $((a))$. Då nu $((a)) \leq (n+1)\,\delta$, är

$$\sum_{\nu=1}^{m+p} a_\nu < (n+1)\,\delta = n\delta + \delta = n\delta_1 - \delta_2,$$

där δ_2 är ett tal mindre än δ.

Således är

$$\sum_{\nu=m+1}^{m+p} a_\nu < \delta_2 < \delta,$$

och detta huru stort talet p än må väljas. Således

$$((a))^{m+1} < \delta_2 < \delta,$$

då med $((a))^{m-1}$ förstås sammanfattningen af alla element i en grupp, hvilka kvartstå, sedan m gifna element a_1, a_2, ... a_m blifvit frånskilda. Q. e. d.

Omvändt gäller äfven:

»Låt (a) vara en grupp af tal i allmännaste mening (cf. § 1), δ ett godtyckligt litet tal samt låt oss från (a) utbryta vissa element a_1, a_2 ... a_m. Kan det hela talet m väljas så stort, att den grupp, som återstår, icke innehåller δ, så är (a) en grupp $((a))$.«

Om nämligen a_{m+1}, a_{m+2}, ... a_{m+p} äro element i den grupp, som återstår efter frånskiljandet af elementen a_1 ... a_m, så är

$$\sum_{\nu=m+1}^{m+p} a_\nu^t < \delta,$$

huru stort jag än må ha valt talet p. Således är äfven, oberoende af talet p,

$$\sum_{\nu=1}^{m+p} \alpha_\nu = \sum_{\nu=1}^{m} \alpha_\nu + \sum_{\nu=m+1}^{m+p} \alpha_\nu < \sum_{\nu=1}^{m} \alpha_\nu + \delta$$

och följaktligen gruppen (a) en grupp $((a))$. Q. e. d.

5. Ett enkelt exempel skall visa oss, att vi i våra grupper $((a))$ erhållit något annat och mera än de tal, vi förut känna. Vi välja den grupp $\left(1, \frac{1}{\lfloor\nu}\right)$, som utgör sammanfattningen af talen $1, \frac{1}{\lfloor\nu}$; $(\lfloor\nu = 1\ 2\ 3\ \ldots\ \nu)$; $\nu = 1, 2, 3 \ldots$, hvilken representerar basen e i det naturliga logaritmsystemet, en af den matematiska analysens tidigast kända och mest beaktade af sådana grupper $((a))$, som ej äro lika med något vare sig helt eller brutet tal.

Man har

$$\sum_{\nu=p}^{p+q} \frac{1}{\lfloor \nu} = \frac{1}{\lfloor p} \left(1 + \frac{1}{p+1} \cdots + \frac{1}{(p+1)\ldots(p+q)}\right)$$

$$= \frac{1}{p} + \frac{1}{\lfloor p+1} \left(1 + \frac{1}{p+2} + \frac{1}{(p+2)(p+3)} + \cdots\right.$$

$$\left. + \frac{1}{(p+2)(p+3)\ldots(p+q)}\right) < \frac{1}{\lfloor p} + \frac{1}{\lfloor p+1}$$

$$\cdot \left(1 + \frac{1}{p+2} + \frac{1}{(p+2)^2} + \cdots + \frac{1}{(p+2)^{q-1}}\right)$$

$$= \frac{1}{\lfloor p} + \frac{1}{\lfloor p+1} \cdot \frac{1 - \frac{1}{(p+2)^q}}{1 - \frac{1}{p+2}} < \frac{1}{p} + \frac{1}{p} \cdot \frac{p+2}{(p+1)^2}$$

$$= \frac{1}{\lfloor p} + \frac{1}{\lfloor p} \left(\frac{1}{p+1} + \frac{1}{(p+1)^2}\right).$$

Om således q är ett helt tal, huru stort som hälst, är

$$\sum_{\nu=1}^{1+q} \frac{1}{\lfloor \nu} < 1 + \tfrac{3}{4},$$

hvaraf åter följer

$$2 < 1 + \sum_{\nu=1}^{1+q} \frac{1}{\lfloor \nu} < 2 + \tfrac{3}{4}$$

samt att gruppen $\left(1, \frac{1}{\lfloor \nu}\right)$ är en grupp $((a))$ eller $\left(\left(1, \frac{1}{\lfloor \nu}\right)\right)$, hvilken har egenskapen

$$2 < \left(\left(1, \frac{1}{\lfloor \nu}\right)\right) \leq 2 + \tfrac{3}{4}$$

och således icke kan vara lika med något helt tal.

Gruppen $\left(\left(1, \frac{1}{\lfloor \nu}\right)\right)$ kan icke heller vara lika med något brutet tal, låt vara $\frac{p_1}{p}$ $(p > 1)$. Vi ha i föregående § visat,

att om δ är ett tal, huru litet som hälst, kan man alltid fastställa ett annat tal $\delta_1 < \delta$ samt ett helt tal n, det förra δ_1 närbeläget δ och det senare n så stort, att

$$\left(\left(1, \frac{1}{\lfloor \nu}\right)\right)^{n+1} < \delta_1 < \delta.$$

Häraf följer åter, att summan af ett antal element huru stort som hälst uti gruppen $\left(\left(1, \frac{1}{\lfloor \nu}\right)\right)$ är mindre än

$$1 + \frac{1}{\lfloor 1} + \ldots + \frac{1}{\lfloor n} + \delta_1,$$

och att således

$$\left(\left(1, \frac{1}{\lfloor \nu}\right)\right) < 1 + \frac{1}{\lfloor 1} + \ldots + \frac{1}{\lfloor n} + \delta$$

samt att, när

$$\frac{p_1}{p} = \left(\left(1, \frac{1}{\lfloor \nu}\right)\right)$$

$$\frac{p_1}{p} < 1 + \frac{1}{\lfloor 1} + \frac{1}{\lfloor 2} + \ldots + \frac{1}{\lfloor n} + \delta.$$

Sätta vi nu $n = p + q$, erhålles

$$\frac{p_1}{p} < 1 + \frac{1}{\lfloor 1} + \frac{1}{\lfloor 2} + \ldots + \frac{1}{\lfloor p-1} + \sum_{\nu=p}^{p+q} \frac{1}{\lfloor \nu} + \delta$$

$$< 1 + \frac{1}{\lfloor 1} + \ldots + \frac{1}{\lfloor p-1} + \frac{1}{\lfloor p} \left(1 + \frac{1}{p+1} + \frac{1}{(p+1)^2}\right) - \delta,$$

hvaraf åter följer

$$p_1 \lfloor p - 1 < P + 1 + \frac{1}{p+1} + \frac{1}{(p+1)^2} - \delta \lfloor p,$$

där P är ett helt tal gifvet i och med talet p samt δ är ett tal huru litet som hälst. Enär P och p $(p > 1)$ äro gifna tal, är således

$$p_1 \lfloor p-1 < P+1+\frac{1}{p+1}+\frac{1}{(p+1)^2}.$$

Å andra sidan innehåller gruppen $\left(\left(1, \frac{1}{\lfloor \nu}\right)\right)$ talet

$$1+\sum_{\nu=1}^{n} \frac{1}{\nu},$$

och om således $\frac{p_1}{p} = \left(\left(1, \frac{1}{\lfloor \nu}\right)\right)$, skulle häraf följa

$$\frac{p_1}{p} > 1+\sum_{\nu=1}^{n} \frac{1}{\lfloor \nu} = 1+\frac{1}{\lfloor 1}+\frac{1}{\lfloor 2}+\ldots+\frac{1}{\lfloor p-1}$$

$$+\sum_{\nu=p}^{p+q} \frac{1}{\lfloor \nu} > 1+\frac{1}{\lfloor 1}+\frac{1}{\lfloor 2}+\ldots+\frac{1}{\lfloor p-1}+\frac{1}{\lfloor p}$$

och således

$$p_1 \lfloor p-1 > P+1$$

samt

$$P+1 < p_1 \lfloor p-1 < P+1+\frac{1}{p+1}+\frac{1}{(p+1)^2}.$$

Talen P och $p_1 \lfloor p-1$ äro hela tal och $\frac{1}{p+1}+\frac{1}{(p+1)^2}$ ett egentligt bråk, och ett helt tal skulle härmed vara beläget mellan ett annat helt tal och summan af detta tal och ett egentligt bråk. Då detta icke kan inträffa, och $\left(\left(1, \frac{1}{\lfloor \nu}\right)\right)$ icke heller är lika med något helt tal, ha vi således i gruppen

$$e = \left(\left(1, \frac{1}{\lfloor \nu}\right)\right)$$

erhållit en grupp $((a))$, som ej är lika med något vare sig helt eller brutet tal. Q. e. d.

»Begreppet grupp är allmännare än begreppet helt eller brutet tal. Det finnes grupper, som icke äro lika med något helt eller brutet tal.«

6. Denna sats är åter endast ett speciellt fall af en allmännare sats: »Låt $A < B$ vara tvänne tal hvilka som hälst och saledes äfven hvarandra huru närbelägna som hälst. Det finnes alltid en grupp $((a))$, hvilken icke är lika med något vare sig helt eller brutet tal och hvilken är belägen mellan A och B, $A < ((a)) < B$.«[1]

Denna sats inbegripes i den beviskedja, hvilken Weierstrass infört för härledningen af en annan sats, hvilken åter utgör den ofrankomliga grunden för hvarje bestandande funktionsteori. Satsen, hvilken först af Weierstrass blifvit skarpt formulerad och samtidigt strängt bevisad, lyder:

»Låt A och B vara tvänne tal eller grupper, af hvilka $A < B$. Antag vidare, att inom intervallet $[A < B]$ finnas angifna ett obegränsadt antal[2] med hvarandra olika tal eller grupper $((a))$. Intervallet $[A < B]$ inbegriper i så fall atminstone en grupp $((c))$, hvilken är ett gränsställe till grupperna $((a))$.«

»Med gränsställe $((c))$ förstas da en sadan grupp, att om densamma inneslutes mellan tvänne tal, hvilka man närmar hvarandra, huru mycket som hälst, sa finnes alltid mellan dem ett obegränsadt antal af de gifna grupperna $((a))$.«

Som exempel må antecknas

[1] Cf. Georg Cantor, »Ueber eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen«, § 2, pag. 260. Journal f. d. reine und angew. Math., Bd. 77. — »Ueber uendliche lineare Punktmannigfaltigkeiten«, pag. 5 seq., Math. Ann., Bd. 15. — »Sur une propriété du système de tous les nombres algébriques réels«, § 2, pag. 308 seq, Acta Math., T. 2.

Henri Poincaré, »Ueber transfinite Zahlen«, pag. 45 seq. Sechs Vorträge aus d. reinen Math. u. math. Physik. Fünfter Vortrag.

[2] = flere med hvarandra olika tal eller grupper än hvilket huru stort helt tal som hälst = oändligt många med hvarandra olika tal eller grupper.

$$A = 1,\ B = 2$$
$$((a)) = 1 + \frac{1}{2^n};\ n = 1, 2, 3 \ldots$$
$$((c)) = 1.$$
$$A = 2,\ B = 3$$
$$((a)) = 1 + \sum_{\nu=1}^{n} \frac{1}{|\nu};\ n = 1, 2, 3 \ldots$$
$$((c)) = \text{gruppen } e \text{ (cf. § 5)}$$

Den Weierstrassiska beviskedjan inbegriper likaledes Georg Cantors berömda sats.[1]

Hat man eine einfach unendliche Reihe $\omega_1, \omega_2, \ldots \omega_\nu, \ldots$ von reellen, ungleichen Zahlen, die nach irgend einem Gesetzt fortschreiten, so lässt sich in jedem vorgegebenen Intervalle $(\alpha \ldots \beta)$ eine Zahl η (und folglich lassen sich deren unendlich viele) angeben, welche nicht in jener Reihe (als Glied derselben) vorkommt.

§ 7. Vi ha i kapitel II visat, huruledes ett antal gifna tal, hela eller brutna, alltid kunna öfverföras i en enhetlig, för dem alla gemensam form (multipler af samma del af enheten). Det var härigenom en jämförelse mellan sådana tal blef möjlig och fragan om likhet eller olikhet, större och mindre kunde afgöras. Vi vilja nu undersöka, på hvad sätt samma problem kan och bör behandlas för våra grupper $((a))$. Vi skola härvid finna, huru i själfva verket äfven samtliga grupper $((a))$ kunna ombildas till en gemensam, enhetlig form.

Vi ha sett (§ 4), att om δ är ett tal, hvilket innehålles i gruppen $((a))$, så finnes alltid ett helt tal n sådant, att

$$n\delta < ((a)) \leq (n+1)\delta.$$

Låt oss att börja med antaga, att gruppen $((a))$ icke

[1] l. c.

innehåller enheten 1. Låt m vara ett helt tal, större än 1. I talföljden $\frac{1}{m_1}, \frac{1}{m_2}, \frac{1}{m_3}, \ldots$ är hvarje efterföljande tal mindre än hvarje föregående. Genom att välja det hela talet γ tillräckligt stort kan man vidare ernå, att $\frac{1}{m^\lambda}$ blir mindre än hvilket angifvet tal huru litet som hälst. Det finnes således ett första tal $\frac{1}{m^\lambda}$, hvilket innehålles i $((a))$. Således finnes också ett tal n_λ (§ 4) sådant, att

$$\frac{n_\lambda}{m^\lambda} < ((a)) \leq \frac{n_\lambda + 1}{m^\lambda}.$$

Då nu $\frac{1}{m^{\lambda+1}} > \frac{1}{m^{\lambda+2}} > \frac{1}{m^{\lambda+3}} > \ldots$ alla innehållas i $((a))$, erhåller man ett obegränsadt antal olikheter

$$\frac{n_\mu}{m^\mu} < ((a)) \leq \frac{n_\mu + 1}{m^\mu}; \quad \mu = \lambda, \lambda+1, \lambda+2, \ldots$$

Af desamma följer

$$\frac{m \cdot n_\mu}{m^{\mu+1}} < ((a)) \leq \frac{m \cdot n_\mu + m}{m^{\mu+1}}$$

$$\frac{n_{\mu+1}}{m^{\mu+1}} < ((a)) \leq \frac{n_{\mu+1} + 1}{m^{\mu+1}}$$

och således

$$m \cdot n_\mu < n_{\mu+1} + 1$$
$$n_{\mu-1} < m \cdot n_\mu + m$$

eller

$$m \cdot n_\mu \leq n_{\mu+1} \leq m \cdot n_\mu + m_1,$$

där m_1 är det hela tal, som närmast föregår m.

Det kan emellertid icke inträffa, att

$$m \cdot n_\mu = n_{\mu+1}$$

från och med en viss index μ. Ty om detta vore fallet, hade man från och med detta μ

$$m \cdot n_\mu = n_{\mu+1}$$
$$m^2 \cdot n_\mu = m \cdot n_{\mu+1} = n_{\mu+2}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$m^{\mu'} \cdot n_\mu = n_{\mu+\mu'}.$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

Å andra sidan kan man alltid, enär $\frac{n_\mu}{m^\mu} < ((a))$, välja μ' så stor, att

$$\frac{n_\mu}{m^\mu} + \frac{1}{m^{\mu+\mu'}} < ((a))$$

eller

$$\frac{m^{\mu'} \cdot n_\mu + 1}{m^{\mu+\mu'}} < ((a)),$$

hvaraf åter följer

$$m^{\mu'} \cdot n_\mu + 1 < n_{\mu+\mu'} + 1$$

eller

$$m^{\mu'} \cdot n_\mu < n_{\mu+\mu'},$$

hvilket strider mot antagandet $m \cdot n_\mu = n_{\mu+1}$ från och med en viss index μ.

Låt oss nu införa en följd af hela tal h_μ; $\mu = \lambda, \lambda+1, \lambda+2 \ldots$, alla mindre än m samt definierade genom likheterna

$$h_\lambda = n_\lambda$$
$$n_{\mu+1} = m \cdot n_\mu + h_{\mu+1};\ \mu = \lambda, \lambda+1, \lambda+2, \ldots,$$

bland hvilka saknas såväl alla tal h_μ, där μ föregår λ, som äfven alla tal $h_{\mu+1}$, för hvilka $n_{\mu+1} = m \cdot n_\mu$. Af denna definition på talen h_μ följer

$$\frac{n_{\lambda+1}}{m^{\lambda+1}} = \frac{n_\lambda}{m^\lambda} + \frac{h_{\lambda+1}}{m^{\lambda+1}} = \frac{h_\lambda}{m^\lambda} + \frac{h_{\lambda+1}}{m^{\lambda+1}}$$

$$\frac{n_{\lambda+2}}{m^{\lambda+2}} = \frac{n_{\lambda+1}}{m^{\lambda+1}} + \frac{h_{\lambda+2}}{m^{\lambda+2}} = \frac{h_\lambda}{m^\lambda} + \frac{h_{\lambda+1}}{m^{\lambda+1}} + \frac{h_{\lambda+2}}{m^{\lambda+2}}$$

och således

$$\frac{n_\mu}{m^\mu} = \frac{h_\lambda}{m^\lambda} + \frac{h_{\lambda+1}}{m^{\lambda+1}} + \ldots + \frac{h_\mu}{m^\mu}$$

$$\mu = 1, 2, 3 \ldots$$

Bilda vi nu en grupp af talen $\frac{h_\nu}{m^\nu}$, ordnade i följden $\nu = 1, 2, 3 \ldots$, hvilken vi skrifva $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$, så är

$$((a)) = \left(\left(\frac{h_\nu}{m^\nu}\right)\right).$$

Lat nämligen $\frac{p}{q}$ vara ett egentligt brak, som innehalles i $((a))$. Man kan uppenbarligen välja ett helt tal μ så stort, att $\frac{p}{q} + \frac{1}{m^\mu} < ((a))$. Å andra sidan är

$$((a)) \leq \frac{n_\mu}{m^\mu} + \frac{1}{m^\mu}$$

och således

$$\frac{p}{q} < \frac{n_\mu}{m^\mu} = \frac{h_\lambda}{m^\lambda} + \frac{h_{\lambda+1}}{m^{\lambda+1}} + \ldots + \frac{h_\mu}{m^\mu} < \left(\left(\frac{h_\nu}{m^\nu}\right)\right).$$

$\frac{p}{q}$ innehålles i $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$ likaväl som i $((a))$.

Om man åter förutsätter, att $\frac{p}{q}$ innehålles i $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$, kan man alltid välja μ så stor, att

$$\frac{p}{q} < \frac{h_\lambda}{m^\lambda} + \frac{h_{\lambda+1}}{m^{\lambda+1}} + \ldots + \frac{h_\mu}{m^\mu} = \frac{n_\mu}{m^\mu}$$

och att således — enär

$$\frac{n_\mu}{m^\mu} < ((a)) \text{ —}$$

$\frac{p}{q}$ innehålles i $((a))$.

Hvarje tal, som innehålles i $((a))$, innehålles äfven i $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$ och vice versa.

Således är

$$((a)) = \left(\left(\frac{h_\nu}{m^\nu}\right)\right).$$

Q. e. d.

Vi ha förutsatt, att gruppen $((a))$ icke innehöll talet 1. Skulle åter detta vara fallet eller $((a))$ innehålla ett största helt tal N, ha vi endast att till gruppen $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$ som element tillfoga talet N.

Vi anteckna till slut, att problemet om jämförelse mellan tvänne olika sammansatta grupper, hvilka ingendera innehålla talet 1, numera erhållit den förenkling, att tvänne grupper $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$ alltid och endast i det fall äro lika hvarandra, att mot samma ν i båda alltid svarar samma h_ν, eller, att om för ett visst ν i den ena gruppen saknas ett motsvarigt h_ν, detsamma är fallet i den andra gruppen, och att om likhet icke äger rum, man endast har att i båda grupperna genomlöpa följden h_1, h_2, h_3 ..., till dess man träffar mot samma ν svarande olika h_ν (den grupp är den större, där h_ν är det större) eller man träffar ett h_ν, som icke har någon motsvarighet i den andra gruppen. Skulle åter vår grupp innehålla talet 1, är den grupp den större, som har till element det större hela talet.

Vi ha lyckats transformera gruppen $((a))$, inom hvilken ingen bestämd ordning mellan elementen var förutsatt, i en ny grupp, där elementen äro entydigt tillordnade talföljden $\nu = 1, 2, 3 \ldots$, så att till hvarje tal är tillordnadt endast ett eller intet element h_ν och att det sista icke kan inträffa för alla ν från och med ett visst gifvet. Bildningslagen för talen h_ν och härmed för gruppen $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$ förutsätter ett obegränsadt antal operationer. Detta är tydligen ofrånkomligt, för såvidt transformationen af $((a))$ i en en-

hetligt bestämd grupp $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$ skall omfatta de allmännast möjliga grupper $((a))$.

Om man inför $m = 10$, erhåller man den bekanta framställningen genom oändligt decimalbråk.

Problemet om en enhetlig framställning af gruppen $((a))$ kan, man ser det lätt, lösas på många olika sätt. Ett af de mest kända är framställningen genom kedjebråk. Vi ha emellertid valt att i det föregående liksom i närmast följande båda paragrafer, där vår uppgift blir att härleda de djupt liggande satser af WEIERSTRASS och CANTOR, hvilka vi i § 6 uttalat, följa WEIERSTRASS' eget bevisningssätt samt begagna den form han valt för framställning af gruppen $((a))$.

Uppgiften att vinna en enhetlig framställning af olika matematiska kombinationer eller, hvilket är detsamma, talkombinationer, hvilken har den största möjliga omfattning, sammanfaller på det närmeste med matematikens uppgift öfverhufvud. Med densamma står eller faller matematiken som vetenskap. En liknande uppgift går för öfrigt igen inom hvarje verklig vetenskap, uppgiften att under enhetlig synpunkt sammanfatta så många som möjligt af vetandets skilda data.

§ 8. Vi återkomma nu till följande sats, hvilken vi uttalat i § 6.

»Låt $A < B$ vara tvänne tal, hvarandra hur närbelägna som hälst. Det finnes alltid en grupp $((a))$ belägen mellan A och B, $A < ((a)) < B$, hvilken icke är lika med något vare sig helt eller brutet tal.«

Bevismetoden sammanfaller så godt som helt och hållet med Weierstrass' bevismetod vid härledningen af satsen beträffande gränsställen (cf. § 6), men vår framställning

vinner i öfverskådlighet, om vi först härleda ofvan anförda satser.

Låt oss med $[a_\nu;\ \nu = 1, 2, 3 \ldots]$ beteckna sammanfattningen af alla hela och brutna tal tillordnade följden af alla hela tal 1, 2, 3 ..., så att mot hvarje helt tal ν svarar ett och endast ett helt eller brutet tal a_ν samt omvändt, mot hvarje tal a_ν ett och endast ett tal ν (cf. Kapitel II, § 7).

Inom hvarje intervall $[A < B]$, det må vara huru litet som hälst, finnes alltid ett obegränsadt antal, på detta sätt definierade tal a_ν (Kap. II, § 7). Låt oss utgå från ett visst dylikt intervall $[A < B]$ och låt oss med a_{θ_1} beteckna det tal a_ν med lägsta index $\nu = \theta_1$, hvilket inbegripes inom intervallet $[A < B]$ eller med andra ord det första af vara tal a_ν, hvilket inbegripes inom detta intervall. Man har således

$$A \leq a_{\theta_1} < B;\ \theta_1 \geq 1,$$

hvarvid — om $\theta_1 > 1$ — intet af talen a_1, a_2 ... $a_{\theta_1 - 1}$ tillhör intervallet $[A < B]$. Låt oss härefter dela intervallet $[A < B]$ i $m \geq 3$ lika stora intervall

$$\left[A < A + \frac{1}{m}(B-A)\right], \left[A + \frac{1}{m}(B-A) < A + \frac{2}{m}(B-A)\right] \ldots$$
$$\left[A + \frac{m-1}{m}(B-A) < B\right].$$

Bland dessa intervall finnes nödvändigt ett första

$$\left[A + \frac{n_1}{m}(B-A) < A + \frac{n_1+1}{m}(B-A)\right];\ 0 < n_1 < m-1,^1$$

[1]) Oaktadt tecknet 0 ännu icke blifvit definieradt, begagna vi oss dock här för korthetens skull af beteckningen $0 \leq n_1$; $0 < n_1$ betyder då, att n_1 är ett helt tal; $0 = n_1$ betyder, att n_1 icke förekommer och att således $A + \frac{n_1}{m}(B-A)$ i detta fall är A samt $A + \frac{n_1+1}{m}(B-A)$ är $A + \frac{1}{m}(B-A)$ och att n_2 etc. icke förekommer.

hvilket icke inbegriper a_{θ_1} och således icke heller något af talen $a_1, a_2, \ldots a_{\theta_1}$.

Låt oss nu med a_{θ_2} förstå det första af våra tal a_ν, hvilket tillhör detta intervall och för hvilket således

$$A + \frac{n_1}{m}(B-A) \leq a_{\theta_2} \leq A + \frac{n_1+1}{m}(B-A);$$

$$0 \leq n_1 \leq m-1;\ \theta_2 > \theta_1$$

samt dela äfven det så erhållna intervallet i m lika delar. Bland de då erhållna nya delintervallen finnes nödvändigt ett första

$$\left[A + \frac{n_2}{m_2}(B-A) < A + \frac{n_2+1}{m_2}(B-A)\right];$$

$$m \cdot n_1 \leq n_2 \leq m \cdot n_1 + m - 1,$$

hvilket icke inbegriper a_{θ_2} och således intet af talen a_1, $a_2, \ldots a_{\theta_1}, a_{\theta_1+1}, \ldots a_{\theta_2}$.

Om vi fortsätta detta förfarande, erhalla vi ett intervall

$$\left[A + \frac{n_\mu}{m^\mu}(B-A) < A + \frac{n_\mu+1}{m^\mu}(B-A)\right];$$

$$m \cdot n_{\mu-1} \leq n_\mu \leq m \cdot n_{\mu-1} - m - 1,$$

hvilket icke inbegriper något af talen

$$a_1, a_2, \ldots a_{\theta_1}, a_{\theta_2}, \ldots a_{\theta_\mu}\ (1 \leq \theta_1 < \theta_2 \ldots < \theta_\mu),$$

och där $a_{\theta_{\mu-1}}$ är det första tal a_ν, som inbegripes i intervallet. Hvarje nytt intervall är en del af det närmast föregående. Det kan dock icke inträffa, att från och med ett gifvet μ, låt vara $\mu = \lambda$, en af de båda likheterna $n_{\mu+1} = m \cdot n_\mu$ eller $n_{\mu+1} = m \cdot n_\mu + m - 1$ städse äger rum. Om så vore fallet, vore nämligen antingen $A + \frac{n_\mu}{m^\mu}(B-A)$ eller $A + \frac{n_\mu+1}{m^\mu}(B-A)$ från och med $\mu = \lambda$ alltid samma tal, låt oss sätta C. Talet C är åter ett af våra tal a_ν, låt

vara $a_{\bar{\nu}}$. Då nu alltid $\mu \leq \theta_\mu$ och intet af talen $a_1, a_2, \ldots$ a_{θ_μ} tillhör intervallet

$$\left[A + \frac{n_\mu}{m^\mu}(B-A) \cdot A + \frac{n_\mu + 1}{m^\mu}(B-A)\right],$$

följer, att C icke kan tillhöra ett sadant intervall, så snart $\mu > \bar{\nu}$. Således är $\frac{n_\mu}{m^\mu} > \frac{n_\lambda}{m^\lambda}$, så snart μ väljes tillräckligt stor. Ingendera af likheterna $n_{\mu+1} = m \cdot n_\mu$ eller $n_{\mu+1} = m \cdot n_\mu + m - 1$ kan således äga rum från och med ett visst tal μ.

Vi återga nu till vår bevismetod i föregaende paragraf.

Vi sätta

$$n_1 = h_1 \leq m - 1$$
$$n_{\mu+1} = m \cdot n_\mu + h_{\mu+1};\ h_{\mu+1} \leq m - 1;\ \mu = 1, 2, 3 \ldots$$

där, som vi visat, talet $h_{\mu+1}$ endast saknas, när

$$n_{\mu-1} = n \cdot n_\mu.$$

Hvarken denna likhet eller likheten

$$n_{\mu+1}' = m - 1$$

kunna, som vi visat, äga rum från och med ett visst tal μ.

Vi bilda nu gruppen $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$; $\nu = 1, 2, 3 \ldots$

Gruppen

$$A + \left(\left(\frac{h_\nu}{m^\nu}\right)\right)(B-A) = \left(\left(A \cdot \left(\left(\frac{h_\nu}{m^\nu}(B-A)\right)\right)\right)^1\right.$$

»innehåller« då alla tal

$$A + \left(\frac{h_1}{m} + \frac{h_2}{m_2} + \ldots \frac{h_\mu}{m^\mu}\right)(B-A) = A + \frac{n_\mu}{m^\mu}(B-A)$$

[1]) Härmed förstås en grupp $((a))$, hvars enskilda element äro talet A samt talen $\frac{h_\nu}{m^\nu}(B-A)$; $\nu = 1, 2, 3 \ldots$

och man har således

$$A+\left(\left(\frac{h_\nu}{m^\nu}\right)\right)(B-A)>A+\frac{n_\mu}{m^\mu}(B-A),$$

huru stor man än må välja index μ. Vidare och då likheten

$$\frac{h_{\mu+1}}{m^{\mu+1}}+\frac{h_{\mu+2}}{m^{\mu+2}}+\ldots+\frac{h_{\mu+\mu'}}{m^{\mu+\mu'}}$$
$$=\frac{m-1}{m^{\mu+1}}\left(1+\frac{1}{m}+\ldots+\frac{1}{m^{\mu'-1}}\right)$$

icke kan äga rum för alla tal μ', som äro större än ett visst gifvet tal, och det således, om μ' väljes tillräckligt stor, alltid finnes ett tal $\bar{\mu}<\mu'$ så stort, att $h_{\mu+\bar{\mu}}<m-1$, så har man för sådana värden på μ'

$$\frac{h_{\mu+1}}{m^{\mu+1}}+\frac{h_{\mu+2}}{m^{\mu+2}}+\ldots+\frac{h_{\mu+\mu'}}{m^{\mu+\mu'}}<\frac{m-1}{m^{\mu+1}}\left(1+\frac{1}{m}+\ldots+\frac{1}{m^{\mu'-1}}\right)$$
$$<\frac{m-1}{m^{\mu+1}}\left(1+\frac{1}{m}+\ldots+\frac{1}{m^{\mu'-1}}\right)-\frac{1}{m^{\mu+\bar{\mu}}}$$
$$=\frac{1}{m^\mu}\left(1-\frac{1}{m^{\mu'}}\right)-\frac{1}{m^{\mu+\bar{\mu}}}$$
$$<\frac{1}{m^\mu}-\frac{1}{m^{\mu+\bar{\mu}}}$$

och således, hur stor man än väljer μ'

$$\frac{n_\mu}{m^\mu}+\frac{h_{\mu+1}}{m^{\mu+1}}+\ldots+\frac{h_{\mu+\mu'}}{m^{\mu+\mu'}}<\frac{n_\mu+1}{m^\mu}-\frac{1}{m^{\mu+\bar{\mu}}}$$

hvaraf åter följer

$$\left(\left(\frac{h_\nu}{m^\nu}\right)\right)<\frac{n_\mu+1}{m^\mu}$$

samt härmed

$$A+\frac{n_\mu}{m^\mu}(B-A)<A+\left(\left(\frac{h_\nu}{m^\nu}\right)\right)(B-A)<A+\frac{n_\mu+1}{m^\mu}(B-A)$$

hvilket äger rum, huru stort man än väljer talet μ.

Häraf följer åter, att talet

$$A+\left(\left(\frac{h_\nu}{m^\nu}\right)\right)(B-A)$$

icke kan vara lika med något helt eller brutet tal a_ν, ty om man väljer $\mu > \nu$, ligga samtliga tal a_ν, hvilkas index är mindre eller lika med θ_μ, som åter är större eller lika med μ, utanföre intervallet

$$\left[A+\frac{n_\mu}{m^\mu}(B-A) < A+\frac{n_\mu+1}{m^\mu}(B-A)\right] \text{ Q. e. d.}$$

Vi ha härmed bevisat tillvaron af en grupp $A+\left(\left(\frac{h_\nu}{m^\nu}\right)\right)(B-A)$, hvilken utan att vara lika med något helt eller brutet tal, är belägen mellan tvänne gifna hvarandra huru närbelägna tal som hälst $A < B$. Gången af vår bevisning har varit, att vi aritmetiskt framställt talet $A+\left(\left(\frac{h_\nu}{m^\nu}\right)\right)(B-A)$. Ett karakteristiskt drag för den Weierstrassiska funktionsteorien är, som vi öfver allt skola finna, att Weierstrass alltid söker föra beviset för tillvaron af det ena eller andra matematiska objektet så, att han för detsamma härleder en bestämd aritmetisk framställning.

§ 9. Vi öfvergå nu till den Weierstrassiska satsen (§ 6).

Vi ha sett, att med gränsställe förstås en grupp sådan, att om densamma instänges mellan tvänne gränser (= tvänne andra grupper), hvilka flyttas hvarandra huru nära som hälst, så finnes alltid mellan dem ett obegränsadt antal andra grupper.

Låt nu A och B, där $A < B$, vara tvänne grupper $((a))$ och antag, att mellan dem finnes ett obegränsadt antal andra grupper $((a))$. Dela som i föregående paragraf intervallet $[A < B]$ uti m $(m \geq 2)$ lika delar

$$\left[A < A + \frac{1}{m}(B-A)\right], \quad \left[A + \frac{1}{m}(B-A) < A + \frac{2}{m}(B-A)\right],$$
$$\ldots \left[A + \frac{m-1}{m}(B-A) < B\right],$$

Det finnes då ett första intervall

$$\left[A < A + \frac{1}{m}(B-A)\right]$$

eller

$$\left[A + \frac{n_1}{m}(B-A) < A + \frac{n_1+1}{m}(B-A)\right]; \quad 1 \leq n_1 \leq m-1,$$

hvilket inbegriper ett obegränsadt antal grupper $((a))$. Om vi ytterligare dela detta intervall i m delar, finns på samma sätt bland då erhållna nya intervall ett första

$$\left[A + \frac{n_2}{m^2}(B-A) < A + \frac{n_2+1}{m^2}(B-A)\right];$$
$$m \cdot n_1 \leq n_2 \leq m \cdot n_1 + m - 1,$$

hvilket inbegriper ett obegränsadt antal grupper $((a))$. Genom att fortsätta på samma sätt erhålla vi en rad intervall

$$\left[A + \frac{n_\mu}{m^\mu}(B-A) < A + \frac{n_\mu+1}{m^\mu}(B-A)\right];$$
$$m \cdot n_\mu \leq n_{\mu+1} \leq m \cdot n_\mu + m - 1; \; \mu = 1, 2, 3 \ldots,$$

om hvilka gäller, att hvarje efterföljande är en del af det föregaende, samt hvilka alla inbegripa ett obegränsadt antal grupper $((a))$. Talet

$$\left(A + \left(\left(\frac{h_\nu}{m^\nu}\right)\right)(B-A)\right),$$

där

$$h_{\mu-1} + m \cdot n_\mu = n_{\mu-1}$$

inbegripes i samtliga dessa intervall. Då desamma genom att öka talet μ kunna minskas under hvarje gräns, så

följer, att talet $A + \left(\left(\begin{matrix} h_\nu \\ m^\nu \end{matrix}\right)\right)(B - A)$ är ett gränsställe, hvarmed vår sats är bevisad.

I olikhet mot hvad fallet var vid användningen i §§ 7 och 8 af Weierstrass' slutledningar, kan här inträffa, att i talet $\left(\left(\begin{matrix} h_\nu \\ m^\nu \end{matrix}\right)\right)$ samtliga brak från och med ett visst ν saknas.

Äfvenså gäller, att den grupp, man erhåller som gränsställe, såväl kan vara lika med som icke lika med ett helt eller brutet tal.

§ 10. Vi öfvergå nu till att uppvisa, huru talens räknelagar kunna bringas att oförändradt gälla för vara grupper $((a))$. Vi ha sett, att dessa grupper äro strängt definierade talobjekt, hvilka på samma gång de omfatta de hela och de brutna talen ha en realitet, hvilken är gifven i och med talens egen. Kunna vi nu härtill visa att talens räknelagar oförandrade gälla för desamma, att man kan räkna med grupper $((a))$ på alldeles samma sätt som med tal, så är härmed adagalagdt, att de äro tal och tal i vidare omfattning än vi förut känna.

Vi börja med begreppet summa af grupper. Detta begrepp framträder så godt som omedelbart. Det är nämligen tydligt, att en grupp, bildad af elementen $a + b$, där a är elementen i en grupp $((a))$ och b elementen i en grupp $((b))$, själf är en grupp $((a + b))$. Det finnes nämligen alltid ett tal l så stort, att det icke innehalles i hvarken $((a))$ eller $((b))$. Häraf följer åter, att talet $2l$ icke kan innehållas i en grupp, bildad af elementen $a + b$ och att således $((a + b))$ är en grupp. I öfverensstämmelse härmed blir definitionen å summa:

Med summan af tvänne grupper $((a))$ och $((b))$ förstå vi gruppen $((a + b))$.

Vi erhålla omedelbart följande satser:

Associationslagen: »$((a)) + [((b)) + ((c))] = [((a)) + ((b))] + ((c))$«

Commutationslagen: »$((a)) + ((b)) = ((b)) + ((a))$« samt:

»Vid summation af ett flertal grupper (a) är summationens ordningsföljd likgiltig.«

Om $((a)) + ((b)) = ((c))$, är $((a)) < ((c))$, $((b)) < ((c))$«.

»Om $((a)) = ((b))$, $((c)) = ((d))$, så är $((a)) + ((c)) = ((b)) + ((d))$.«

»Om $((a)) > ((b))$, $((c)) > ((d))$, så är $((a)) + ((c)) > ((b)) + ((d))$.«

»Om $((a)) = ((b))$ och $((c)) > ((d))$ så är $((a)) + ((c)) > ((b)) + ((d))$.«

»Om $((a)) = ((b))$ och $((a)) + ((c)) > ((b)) + ((d))$, så är $((c)) > ((d))$.«

»Om $((a)) < ((b))$ samt $((a)) + ((c)) > ((b)) + ((d))$ så är $((c)) > ((d))$.«

»Om $((a)) < ((b))$ och $((a)) + ((c)) = ((b)) + ((d))$, så är $((c)) > ((d))$.«

»Om $((a)) = ((b))$ och $((a)) + ((c)) = ((b)) + ((d))$, så är $((c)) = ((d))$.«

Samma lagar, som gälla addition af tal, gälla således oförändrade vid addition af grupper $((a))$.

Häraf följer ater, att vi kunna behandla grupper (A), hvilkas element A äro grupper $((a))$, pa samma sätt som de grupper, där dessa element äro tal.

Låt nämligen (A) vara en grupp, hvars element äro grupper $((a))$. I öfverensstämmelse med § 1 kommer

»ett tal eller en grupp l att innehållas i en grupp, bildad af element $((a))$, om man från gruppen kan frånskilja ett antal element, hvilkas summa är större än l.«

»Elementen $A = ((a))$ bilda en grupp $((A))$, om det finnes ett tall l, hvilket icke innehålles i gruppen.«

»Likhet mellan tvänne grupper $((a))$ och $((b))$, $((a)) = ((b))$, $((b)) = ((a))$, där elementen a och b själfva äro grupper, äger rum, när hvarje tal, som innehålles i den ena gruppen också innehålles i den andra, hvaremot $((a)) > ((b))$, $((b)) < ((a))$, när det finnes ett tal, som innehålles i $((a))$ utan att innehållas i $((b))$.«

Den härledning, genom hvilken vi (§ 7) bringat våra grupper $((a))$ under den enhetliga formen $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$ blir ord för ord densamma, om man i stället för talen a inför grupper $((a))$. »Talvärlden inbegriper icke andra grupper än sådana, som kunna framställas under formen $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$.«

§ 11. Sedan vi infört begreppet summa af grupper, inser man omedelbart, huruledes Georg Cantors sats (§ 6) innehålles i den beviskedja, hvilken vi i § 8 upprullat. Vi ha vid densamma utgått från följden af samtliga hela och brutna tal a_ν, entydigt tillordnade följden af de hela talen 1, 2, 3 ... Om vi nu i stället för talen a_ν införa med hvarandra olika grupper $((a))_\nu$, hvilka bilda en obegränsad följd $((a))_1$, $((a))_2$, $((a))_3$ ... entydigt tillordnade de hela talen 1, 2, 3 ..., så bli alla slutföljder oförändradt desamma, som då talen a_ν voro de hela och brutna talen.

Vi kunna alltid bilda en grupp $((a))$, hvilken icke är lika med någon af grupperna $((a))_\nu$ och som är belägen mellan tvänne hvilka som hälst af desamma.

Grupperna $((a))$ kunna icke samtliga, som fallet var med de hela och brutna talen, ordnas i en följd, entydigt tillordnad talen 1, 2, 3 ...

Man erhåller däremot samtliga grupper $((a)) < 1$, om man i uttrycket $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$ [§ 7] låter talen h_ν hvart och ett för sig antaga alla heltalsvärden $< m-1$ eller försvinna ur uttrycket (utan att alla på en gang eller alla från och med ett visst ν försvinna).

Vi ha genom införande af grupperna $((a))$ träffat på en oändlighet af högre art än den intuitivt gifna oändlighet, hvilken var sammanfattningen af alla hela tal 1, 2, 3 . . . Vi stå inför begreppet kontinuum.

Kontinuet är sammanfattningen af alla grupper $((a))$.

»De olika grupper, som bilda kontinuet, kunna icke samtliga entydigt tillordnas talföljden 1, 2, 3 . . .«

§ 12. Återstår närmast att undersöka, huruvida samma sats, som gällde skillnad mellan tvänne tal (kapitel I, § 6), kan utsträckas att gälla skillnad mellan två grupper. Detta är, som vi lätt skola se, verkligen fallet.

»Om $((b)) > ((a))$, kan man alltid bilda en ny grupp $((c))$ sådan, att $((b)) = ((a)) + ((c))$.«

Antag $((b)) = h' + \left(\left(\frac{h'_\nu}{m^\nu}\right)\right)$ och $((a)) = h + \left(\left(\frac{h_\nu}{m^\nu}\right)\right)$ (cf. § 7), där h' och h äro hela tal, af hvilka enbart h eller båda h' och h kunna saknas, samt $h'_\nu \leq m-1$ och $h_\nu \leq m-1$, och beträffande såväl h'_ν som h_ν gäller, att de för vissa index ν kunna saknas i respektive grupper, men att detta dock icke kan äga rum från och med en viss minsta index.

Enär $((b)) > ((a))$ finnes alltid ett tal l, som innehålles i $((b))$ utan att innehållas i $((a))$. Således är alltid

$$l \geq h + \left(\left(\frac{h_\nu}{m^\nu}\right)\right)$$

och således

$$l > h + \frac{h_1}{m} + \frac{h_2}{m^2} + \ldots + \frac{h_\nu}{m^\nu} = \frac{n_\nu}{m^\nu} \geq \frac{1}{m^\nu}$$

samt om man fastställer λ tillräckligt stor

$$h' + \frac{h'_1}{m} + \frac{h'_2}{m^2} + \ldots + \frac{h'_\nu}{m^\nu} = \frac{n'_\nu}{m^\nu} > l;\ \nu = \lambda,\ \lambda+1,\ \lambda+2, \ldots,$$

hvaraf åter följer

$$\frac{n'_\nu}{m^\nu} > \frac{n_\nu}{m^\nu};\ \nu = \lambda,\ \lambda+1,\ \lambda+2 \text{ etc.} \ldots$$

eller $n'_\nu - 1 \geq n_\nu \geq 1$.

Å andra sidan är

$$1 = \left(\frac{m-1}{m^\nu}\right)$$

och gruppen $((b))$ kan således uttryckas under formen

$$((b)) = \frac{n'_\lambda - 1}{m^\lambda} + \frac{1}{m^\lambda} + \left(\left(\frac{h'_{\lambda+\mu}}{m^{\lambda+\mu}}\right)\right)$$

$$= \frac{n'_\lambda - 1}{m^\lambda} + \left(\left(\frac{m-1}{m^{\lambda+\mu}}\right)\right) + \left(\left(\frac{h'_{\lambda+\mu}}{m^{\lambda+\mu}}\right)\right)$$

$$= \frac{n'_\lambda - 1}{m^\lambda} + \left(\left(\frac{h'_{\lambda+\mu} + m - 1}{m^{\lambda+\mu}}\right)\right);\ \mu = 1,\ 2,\ 3 \ldots$$

Vidare är

$$h'_\nu + m - 1 \geq h_\nu \text{ och } h'_\nu + m - 1 \leq 2(m-1).$$

Låt oss nu införa en serie af likheter

$$n_\lambda + k_\lambda = n'_\lambda - 1$$
$$h_{\lambda+\mu} + k_{\lambda+\mu} = h'_{\lambda+\mu} + m - 1,$$

så erhålla vi uti $\frac{k_\lambda}{m^\lambda} + \left(\left(\frac{k_{\lambda+\mu}}{m^{\lambda+\mu}}\right)\right)$ den sökta gruppen $((c))$.

Ty

$$((b)) = \frac{n'_\lambda - 1}{m^\lambda} + \left(\left(\frac{h'_{\lambda-\mu} + m - 1}{m^{\lambda+\mu}}\right)\right)$$

$$= \frac{n_\lambda}{m^\lambda} + \left(\left(\frac{h_{\lambda-\mu}}{m^{\lambda+\mu}}\right)\right) + \frac{k_\lambda}{m^\lambda} + \left(\left(\frac{k_{\lambda-\mu}}{m^{\lambda-\mu}}\right)\right) = ((a)) + ((c)).$$

Vissa af talen k_λ, $k_{\lambda+1}$ ... kunna saknas i uttrycket för $((c))$, dock icke alla från och med en viss index, ty i så fall skulle $\left(\left(\frac{h'_{\lambda-\mu}}{m^{\lambda-\mu}}\right)\right)$ icke komma att innehalla ett obegränsadt antal element.

Gruppen $\left(\left(\frac{k_{\lambda-\mu}}{m^{\lambda-\mu}}\right)\right)$ är icke en sadan grupp, som vi betecknat med $\left(\left(\frac{h_\nu}{m^\nu}\right)\right)$, men kan i öfverensstämmelse med § 7 alltid förvandlas i vare sig en sådan grupp eller i summan af en dylik och ett helt tal.

Att gruppen (c) är entydigt bestämd följer af § 10.

Låt oss nu undersöka, hvad som bör förstås med produkt af grupper $((a))$. Lat $((a))$ vara en grupp och m ett tal. Da $((a))$ är en grupp, finnes alltid ett tal l sådant, att $\sum_{\nu=1}^{n} a_\nu < l$, huru stort man än ma välja n. Följaktligen är

$$\sum_{\nu=1}^{n} m \cdot a_\nu < m \cdot l,$$

huru stort jag än må välja n. Således den grupp, som bildas af elementen $m \cdot a$ är en grupp $((m \cdot a)) = ((a \cdot m))$.

»Med produkten af talet m och gruppen $((a))$ förstå vi gruppen $((m \cdot a)) = ((a \cdot m))$ och skrifva

$$m \cdot ((a)) = ((a)) \cdot m = ((m \cdot a)) = ((a \cdot m)).«$$

Man erhåller omedelbart

$$»\left[\sum_{\nu=1}^{n} m_\nu\right]\cdot((a)) = \sum_{\nu=1}^{n} m_\nu\,((a)) = \sum_{\nu=1}^{n}((m_\nu\, a)).«$$

Låt oss nu i $((a))$ och $((b))$ ha tvänne olika grupper. Det finnes alltid ett tal l sådant, att

$$\sum_{\nu=1}^{n} a_\nu < l,$$

huru stort jag än må välja talet n. Vidare finnes alltid ett tal m sådant, att $((b)) < m$. Man erhåller således

$$\sum_{\nu=1}^{n} a_\nu\,((b)) = \left[\sum_{\nu=1}^{n} a_\nu\right]\cdot((b)) < m\cdot l.$$

Följaktligen är en grupp bildad af elementen $a\cdot((b)) = ((b))\cdot a$, en grupp $\big((a((b)))\big) = \big(((b))\,a\big)$ och på samma sätt en grupp, bildad af elementen $b\cdot((a)) = ((a))\cdot b$ en grupp $\big(\big(b\,((a))\big)\big) = \big(\big(((a))\,b\big)\big)$.

Låt nu l vara ett tal, som innehålles i gruppen $\big(\big(a\cdot((b))\big)\big)$. Det finnes då alltid ett tal n_1 så stort, att

$$\sum_{\nu=1}^{n_1} a_\nu\,((b)) > l$$

och således också

$$\sum_{\nu=1}^{n_1} a_\nu\,((b)) > l+\delta$$

om δ väljes tillräckligt liten. Å andra sidan finnes alltid ett tal n_2 så stort, att

$$\sum_{\nu=1}^{n_1} a_\nu\,((b))^{n_2+1} < \delta.^{1}$$

[1] Cf. rörande beteckningssättet $((b))^{n_2+1}$ § 4.

Således

$$\sum_{\nu=1}^{n_1} a_\nu \left[\sum_{\mu=1}^{n_2} b_\mu + ((b))^{n_2+1} \right] > l + \delta,$$

hvaraf följer

$$\sum_{\nu=1}^{n_1} a_\nu \left[\sum_{\mu=1}^{n_2} b_\mu \right] > l$$

eller

$$\sum_{\mu=1}^{n_2} b_\mu \sum_{\nu=1}^{n_1} a_\nu > l.$$

Häraf följer åter

$$\sum_{\mu=1}^{n_2} b_\mu ((a)) > l.$$

Talet l, som innehålles i $\big(a((b))\big)$, innehålles alltid samtidigt i $\big(b((a))\big)$, och likheten

$$\Big(\big(a((b))\big)\Big) = \Big(\big(b((a))\big)\Big)$$

äger följaktligen rum.

Definitionen på produkt blir uppenbarligen:

»Med produkten af gruppen $((a))$ och gruppen $((b))$, $((a))\cdot((b))$, förstås gruppen $\Big(\big(a((b))\big)\Big) = \Big(\big(b((a))\big)\Big)$.«

Vi ha redan sett, att commutationslagen

$$\text{»}((a))\cdot((b)) = ((b))\cdot((a))\text{«}$$

äger rum. En enkel betraktelse visar, att detta äfven är fallet med associationslagen eller att

$$\text{»}\big[((a))\cdot((b))\big]\cdot((c)) = ((a))\cdot\big[((b))\cdot((c))\big].\text{«}$$

Man ser äfven, att vid multiplikation af ett begränsadt flertal af grupper $((a))$ är det likgiltigt, i hvilken ordning

multiplikationen företages. Man erhåller alltid till slutresultat en och samma grupp.

Vidare gäller:

»Om $((a)) = ((b))$ och $((c)) \geq ((d))$, så är $((a))\cdot((c)) \geq ((b))\cdot((d))$.«

Om $((a)) > ((b))$ och $((c)) > ((d))$, så är $((a))\cdot((c)) > ((b))\cdot((d))$.

samt omvändt

»Om $((a)) = ((b))$ och $((a))\cdot((c)) > ((b))\cdot((d))$, så är $((c)) > ((d))$.

Om $((a)) = ((b))$ och $((a))\cdot((c)) = ((b))\cdot((d))$, så är $((c)) = ((d))$.

Om $((a)) > ((b))$ och $((a))\cdot((c)) = ((b))\cdot((d))$, så är $((c)) < ((d))$.«

§ 14. Vi ha visat, huru lagarna för addition (§ 10), subtraktion (§ 12) samt multiplikation (§ 13) oförändrade gälla för våra grupper $((a))$. Begreppet division gäller likaledes oförändradt.

»Om $((b))$ och $((a))$ äro tvänne gifna grupper, kan man alltid bilda en ny grupp $((c))$ sådan, att $((b)) = ((a))\cdot((c))$.«

Låt nämligen m vara ett helt tal, som icke innehålles i $((a))$. Det finnes alltid en grupp $((\alpha))$ sådan, att (§ 12)

$$m = ((a)) + ((\alpha))$$

och att således

$$1 = \frac{((a))}{m} + \frac{((\alpha))}{m}.$$

Låt oss nu införa en ny grupp

$$((\beta)) = 1 + \left(\left(\left[\frac{((\alpha))}{m}\right]^{\nu}\right)\right); \quad \nu = 1, 2, 3 \ldots$$

Man har

$$\frac{((\alpha))}{m}\cdot((\beta)) = \left(\left[\frac{((\alpha))}{m}\right]^{\nu}\right); \quad \nu = 1, 2, 3 \ldots$$

och således

$$((\beta)) = 1 + \frac{((\alpha))}{m} \cdot ((\beta))$$

samt om vi å båda sidor om likhetstecknet addera $\frac{((a))}{m}((\beta))$

$$((\beta)) + \frac{((a))}{m}((\beta)) = 1 + \frac{((a))}{m} + \frac{((a))}{m}((\beta)) = 1 + ((\beta)).$$

Häraf följer åter

$$\frac{((a))}{m} \cdot ((\beta)) = 1$$

och således

$$((a))\frac{((b))}{m} \cdot ((\beta)) = ((b))$$

$$((c)) = \frac{((b))}{m} \cdot ((\beta)).$$

Vi ha nu visat, huru samtliga lagar för addition, subtraktion, multiplikation och division, hvilka vi i kapitel I och II härledt för de hela och brutna talen, oförändrade gälla för våra grupper $((a))$.

Grupperna $((a))$, hvilka omfatta de hela och brutna talen och hvilka samtidigt ha en långt vidsträcktare giltighet, äro själfva tal. De kunna med fog nämnas allmänna tal. Den öfliga benämningen för desamma, så snart de icke äro hela eller brutna, är inkommensurabla tal. Då dessa länge och allt intill nyaste tider icke betraktades som begreppsligt gifna objekt, utan antingen som geometriska, genom åskådning vunna storheter eller som symboler, underkastade de verkliga talens räknelagar, kom äfven den alltfortfarande begagnade benämningen irrationella tal att beträffande desamma vinna burskap. I motsats härtill sammanfattades då de hela och brutna talen under den ännu alltjämt antagna benämningen rationella tal.

Sedan vi nu visat, att $((a))$ kan behandlas på alldeles samma sätt som de tal a, hvilka vi förut kallat absoluta tal, inbegripa vi härefter äfven talen $((a))$ under denna benämning. Det allmänna talet i ett talsystem med e som grundenhet kommer således att tecknas $e \cdot ((a))$ och lyder samma räknelager som $e \cdot a$, där a är ett rationelt tal.

§ 15. Det historiskt första problem, hvilket efter årtusendens mer eller mindre medvetet sökande slutligen genom införande af gruppen $((a))$ eller det allmänna absoluta talet erhållit en verklig lösning, torde ha varit den begreppsliga tolkningen af den geometriska erfarenhet, att kvadratens sida och diagonal icke kunna mätas med något gemensamt matt. Vid sökandet efter ett tal x, som uppfyller likheten

$$x^2 = 2,$$

finner man omedelbart, att det icke finnes något dylikt tal (helt eller brutet tal), men den geometriska askadningen ger oss dock för det sökta obekanta, för x, ett visst något. Hvad är detta nagot begreppsligt fattadt? Det föregaende har lärt oss att besvara denna fråga.

Vi bilda aritmetiskt den grupp $((a))$ eller det allmänna tal, hvilket är vårt sökta x.

Vi komma härvid enklast till målet genom samma slutföljder som i § 7.

Låt oss med P förstå ett helt tal, som icke är produkten af ett helt tal större än ett multipliceradt med sig själf. Låt vidare m vara ett helt tal större än ett. Mot talet m^{2^ν}; $\nu = 1, 2, 3 \ldots$ svarar då alltid ett helt tal $n_\nu > 1$ sådant att

$$\left[\frac{n_\nu}{m^\nu}\right]^2 < P < \left[\frac{n_\nu + 1}{m^\nu}\right]^2$$

och således också ett helt tal $n_{\nu+1}$ sådant att

$$\left[\frac{n_\nu+1}{m^\nu+1}\right]^2 < P < \left[\frac{n_\nu+1+1}{m^\nu+1}\right]^2.$$

Man ser, att

$$[n_{\nu-1}+1]^2 > [n_\nu m]^2$$

och således

$$n_{\nu+1}+1 > n_\nu \cdot m$$

$$n_{\nu+1} \geq n_\nu \cdot m.$$

Å andra sidan är äfven

$$n_\nu \cdot m + m > n_{\nu+1}$$

och således är

$$n_\nu \cdot m + m - 1 \geq n_{\nu-1}.$$

Vi införa nu de hela talen h, h_1, h_2, h_3, ... som vi definiera genom likheterna

$$h^2 < P < (h+1)^2$$
$$h_1 + n \cdot m = n_1$$
$$h_2 + n_1 \cdot m = n_2$$
$$\ldots\ldots\ldots\ldots$$
$$h_\nu + n_{\nu-1} \cdot m = n_\nu;\ \nu = 1, 2, 3 \ldots$$
$$\ldots\ldots\ldots\ldots$$

Om dessa tal gäller, att så snart $n_{\nu-1} \cdot m \neq n_\nu$

$$h_\nu \leq m-1$$

samt vidare att så snart $n_\nu \cdot m = n_{\nu+1}$ icke finnes något mot ν svarande tal $h_{\nu+1}$.

Dessutom gäller, att det icke kan inträffa, att från och med ett visst ν, låt vara $\nu = \lambda$

$$n_\nu \cdot m = n_{\nu+1}.$$

Ty emedan

$$\left[\frac{n_\lambda}{m^\lambda}\right]^2 < P < \left[\frac{n_\lambda}{m^\lambda} + \frac{1}{m^{\lambda+\mu}}\right]^2,$$

och man alltid kan välja μ så stor, att

$$\left[\frac{n_\lambda}{m^\lambda}+\frac{1}{m^{\lambda+\mu}}\right]^2 < P \text{ eller } \left[\frac{m^\mu \cdot n_\lambda + 1}{m^{\lambda+\mu}}\right]^2 < P$$

följer

$$m^\mu \cdot n_\lambda + 1 < n_{\lambda+\mu} + 1$$

eller

$$m^\mu \cdot n_\lambda < n_{\lambda+\mu},$$

hvilket åter är oförenligt med $m^\mu \cdot n_\lambda = n_{\lambda+\mu}$ eller att likheten $m \cdot n_\nu = n_{\nu+1}$ äger rum från och med $\nu = \lambda$.

Låt oss nu bilda gruppen

$$S = h + \left(\left(\frac{h_\nu}{m^\nu}\right)\right).$$

Man har, då icke alla h_ν från och med ett visst tal ν kunna i densamma saknas,

$$h + \frac{h_1}{m} + \frac{h_2}{m^2} + \ldots + \frac{h_\nu}{m^\nu} < S.$$

Å andra sidan är

$$h + \frac{h_1}{m} + \frac{h_2}{m^2} + \ldots + \frac{h_\nu}{m^\nu} = \frac{n_\nu}{m^\nu}.$$

Vidare är

$$\frac{h_{\nu+1}}{m^{\nu+1}} + \frac{h_{\nu+2}}{m^{\nu+2}} + \ldots + \frac{h_{\nu+\mu}}{m^{\nu+\mu}}$$

$$< (m-1)\frac{1-\frac{1}{m^\mu}}{1-\frac{1}{m}} \cdot \frac{1}{m^{\nu+1}} < \frac{1}{m^\nu}$$

och således

$$\frac{n_\nu}{m^\nu} < S < \frac{n_\nu + 1}{m^\nu}$$

$$\left[\frac{n_\nu}{m^\nu}\right]^2 < S^2 < \left[\frac{n_\nu + 1}{m^\nu}\right]^2.$$

Man har således

$$P = \left[\frac{n_\nu}{m^\nu}\right]^2 + \delta$$

$$S^2 = \left[\frac{n_\nu}{m^\nu}\right]^2 + \varepsilon,$$

hvarvid beträffande såväl δ som ε gäller

$$\begin{matrix}\delta \\ \varepsilon\end{matrix} \leq \frac{1}{m^\nu}\left[2\frac{n_\nu}{m^\nu} + \frac{1}{m^\nu}\right]$$

$$= \text{skillnaden mellan } \left[\frac{n_\nu + 1}{m^\nu}\right]^2 \text{ och } \left[\frac{n_\nu}{m^\nu}\right]^2.$$

Då åter P är ett gifvet helt tal, kan aldrig

$$\left[\frac{n_\nu}{m^\nu}\right]^2,$$

som är mindre än P och således ej heller $\frac{n_\nu}{m^\nu}$ växa öfver hvarje gräns. Talen δ och ε kunna således genom att tillräckligt öka ν bringas att bli mindre än hvilket huru litet tal som hälst. Hvarje tal, som innehålles i P, innehålles således äfven i S^2 och likheten

$$P = S^2$$

äger således rum.

Det finnes ett allmänt tal

$$x = S,$$

för hvilket

$$x^2 = P.$$

§ 16. Den härledning af det allmänna absoluta talet, hvilken vi gifvit i det föregående, ansluter sig i allt väsentligt till WEIERSTRASS. Den utarbetades af mig under hösten 1877 som inledning till en föreläsning öfver analytiska funktioner. Jag hade härvid tillgång till en programafhandling af E. KOSSAK (Die Elemente der Arithmetik, Berlin

1872), som innehaller ett försök till redogörelse för Weierstrass' teori, hvilken dock alls icke hade dennes gillande. Weierstrass skrifver exempelvis till mig 5. april 1895 i anledning af en annan fråga:

»Sie wissen, wie meine Einleitung in die Functionentheorie von den Herren Kossak und . . . verhunzt worden ist.« Kossak åberopar sig på en föreläsning af Weierstrass vintersemestern 1865—66. För hvar och en, som är förtrogen med det Weierstrassiska framställningssättet och de Weierstrassiska metoderna, står emellertid klart, att han redan tidigare i sina föreläsningar i Berlin meddelat allt väsentligt af sin teori. Så måste i hvarje fall ägt rum vid den föreläsning, hvilken han vintern 1859—60 samt sommaren 1860 höll öfver ämnet »Einleitung in die Analysis«, den förra 5 timmar, den senare 4 timmar i veckan (denna sista gratis), och antagligen redan vid hans föreläsning i sommaren 1857 »Allgemeine Lehrsätze betreffend die Darstellung analytischer Functionen durch convergente Reihen« (publice 1 timme i vecken).[1]

I begreppslig besittning af sin teori var Weierstrass säkerligen redan åren 1841, 1842. Enligt hvad han själf framhållit och hvilket likaså framgår ur de fyra första af honom publicerade afhandlingarna i Werke, Bd. 1, var han vid denna tid färdig med allt det väsentligaste i sin funktionsteori. Själfva basen för den Weierstrassiska funktionsteorien, det oundgängliga villkoret för den matematiska analysens aritmetisering och strängt begreppsliga uppbyggnad från grunden, är emellertid och måste vara fastställande af de båda begreppen allmänt tal och gränsställe, hvilka före Weierstrass icke kunnat tolkas

[1] »Verzeichniss der von Weierstrass an der Universität zu Berlin gehaltenen und angekündigten Vorlesungen.« Werke, Bd. 3, pag. 355, 356.

annat än genom mer eller mindre oklar geometrisk åskådning.

Heine klagar i sin afhandling »Die Elemente der Functionenlehre«,[1] publicerad 1872, samma år som Kossaks skrift, öfver att de Weierstrassiska satserna, ehuru af Weierstrass bevisade »unentbehrliche Fundamentalsätze«, dock endast äro bekanta genom dennes föreläsningar och muntliga meddelanden samt indirekt genom anteckningar efter föreläsningar och att desamma därföre allt fortfarande af många till sin riktighet betviflas. Detta är också skälet, hvarföre Heine, stödd på muntliga meddelanden af Weierstrass själf, samt med verksamt biträde af Cantor offentliggör sitt arbete. Han har emellertid enligt Cantors råd i detsamma så tillvida modifierat Weierstrass' grunddefinition, att han i stället för gruppen $((a))$ utgår från en obegränsad följd af tal, hela eller brutna

$$u_1 < u_2 < u_3 < \ldots,$$

sådan, att differensen mellan ett tal u_n och ett annat u_{n-m}, om blott n väljes tillräckligt stor, oberoende af m, blir mindre än hvilket som hälst tal δ, huru litet detta än må ha blifvit fastställdt. Han benämner denna talföljd Fundamentalreihe. Denna »Fundamentalreihe« definierar nu ett visst något, som blir det nya talet: detta blir en symbol, ett tecken. Också ser sig Heine tvungen uttala: »Ich stelle mich bei der Definition auf den rein formalen Standpunkt, indem ich gewisse greifbare Zeichen Zahlen nenne, so dass die Existenz dieser Zahlen nicht in Frage steht.«

I själfva verket blir Cantor-Heines utgångspunkt

[1] Journal f. d. reine und angew. Math. Bd. 74, pag. 172—188.

$$\lim_{n=\infty} u_n = \lim_{n=\infty} \sum_{\nu=1}^{n} a_\nu = \sum_{\nu=1}^{\infty} a_\nu ,$$

hvilken definition, som vi senare skola finna, vid en framställning af den Weierstrassiska teorien i sin helhet först uppträder, när begreppen allmänt tal med mer än en grundenhet och utredningen af sådana tals väsentliga egenskaper redan ägt rum.

Begreppet gränsställe kommer hos Cantor-Heine att redan vid teoriens ingang postuleras i stället för att som hos Weierstrass bevisas. Serien $\sum_{\nu=1}^{\infty} a_\nu$ är ju ingenting annat än ett gränsställe för talen u_n; $n = 1, 2, 3 \ldots$

Ungefär samtidigt med Cantor-Heine har Méray[1] framställt en teori för det allmänna absoluta talet, hvilket definieras på samma sätt som hos Cantor-Heine. Dessas »Fundamentalreihe« benämnes hos Méray variante. Att Méray lika litet som Cantor-Heine uppfattat det af honom införda »inkommensurabla« talet som ett verkligt tal, framgår af hans reflektion: »Telle est pour nous la nature des nombres incommensurables: ce sont des fictions permettant d'énoncer d'une manière uniforme et plus pittoresque toutes les positions relatives aux invariantes convergentes.«[2]

Samma osäkerhet uppträdde på sin tid beträffande de negativa talen och ännu nära var tid beträffande de komplexa talen. De voro icke verkliga tal, utan en sorts symboler, å hvilka man kunde utföra i det närmaste samma räkneoperationer som å verkliga tal.

I den Weierstrassiska teorien är uppfattningen en annan. Den sammanfattning af tal, som vi kallat det allmänna,

[1] Charles Méray, »Nouveau précis d'analyse infinitésimale«. Paris 1872.
[2] L. c. p. 4.

absoluta talet, är en skapelse, hvilken är gifven i och med det första talbegreppet samt vunnen genom tankens förmåga att tillordna hvarje tal i en talföljd 1, 2, 3, ... ν ... ett nytt tal och uppfatta det objekt, som är sammanfattningen af dessa senare tal, som en realitet, lika bestämd som den obegränsade talföljden 1, 2, 3 ...

Dedekind[1] har infört det allmänna talet från en annan synpunkt än Cantor-Heine-Méray. Hans definition är lånad från den geometriska intuitionen. Han yttrar:

»Zerfallen alle Puncte der Geraden in zwei Classen von der Art, dass jeder Punct der ersten Classe links von jedem Puncte der zweiten Classe liegt, so existiert ein und nur ein Punct, welcher diese Eintheilung aller Puncte in zwei Classen, diese Zerschneidung der Geraden in zwei Stücke hervorbringt;«[2]

samt ytterligare:

»Ich bin ausser Stande, irgend einen Beweis für seine Richtigkeit beizubringen, und Niemand ist dazu im Stande. Die Annahme dieser Eigenschaft der Linie ist nichts als ein Axiom, durch welches wir erst der Linie ihre Stetigkeit zuerkennen, durch welches wir die Stetigkeit in die Linie hineindenken.«[2]

Sedt ur geometrisk synpunkt, kan Dedekinds standpunkt »att det allmänna talets existens icke är annat än ett axiom« motiveras. Ur aritmetisk och begreppslig däremot icke. Dedekinds geometriska definition kan emellertid, som han själf senare framhållit,[3] utan svårighet öfverföras i en rent

[1] »Stetigkeit und irrationale Zahlen«. Braunschweig 1872. 2. Aufl. 1892. Jmfr. dessutom »Was sind und was sollen die Zahlen?« Braunschweig 1888.

[2] L. c. pag. 18.

[3] »Was sind und was sollen die Zahlen?« Vorwort. Harzburg 5 okt. 1887.« Braunschweig 1887, 1893.

aritmetisk. Detta har skett hos Lipschitz i hans fem år efter »Stetigkeit und irrationale Zahlen« utgifna »Lehrbuch der Analysis«.[1] När man följer Dedekind, utgår man emellertid alltid, huru man än må variera sin framställning, från två »Fundamentalreihen« eller »variantes« i stället för som hos Cantor-Heine eller Méray från blott en.

Dedekinds teori har emellertid vunnit större insteg i läroboksliteraturen än någon annan. Förklaringen synes ligga i dess vädjan till geometrisk intuition. Man inlägger inom själfva grunden för aritmetiken en geometrisk åskådning, hvilken är och bör vara för densamma främmande, men vinner å andra sidan den skenbart större lättfattlighet, som geometriskt åskådande synes medföra.

[1] Bd. 1., Bonn 1877.

Færdig fra Trykkeriet den 23. April 1920.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **II**, 6.

EXPERIMENTALUNDERSØGELSER OVER GNIDNINGSELEKTRICITETENS OPRINDELSE. VII

AF

C. CHRISTIANSEN
OG
JOHANNE CHRISTIANSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1919

Om amphotere Stoffers Balloelektricitet.

Fra Institut for alm. Patologi.

Hensigten med at undersøge amphotere Stoffers balloelektriske Forhold var at faa at vide, om disse Stoffers ejendommelige Forhold til Syrer og Baser viste sig ogsaa balloelektrisk. Da C. Christiansen[1] har fundet, at en Elektrolyt, naar den i vandig Opløsning bringes sammen med visse ikke ledende Stoffer, f. Ex. Alkohol, giver et meget stærkt balloelektrisk Udslag, kunde man tænke sig den Mulighed, at der i en Række af f. Eks. Æggehvideopløsninger med samme Concentration Æggehvide, men stigende Mængde Syre eller Base, vilde begynde at optræde særlig stærk Balloelektricitet i Nærheden af det Punkt, hvor Æggehviden var mættet med Syre eller Base, saaledes at der i Opløsningen begynder at optræde hydrolyseret eller fri Syre.

Da disse Punkter paa anden Maade er kendelige, f. Ex. ved Optræden af et Viscositetsmaximum[2] og Reaktion paa fri Syre (Günzburgs Reaktion)[3], var der Mulighed for, at saadanne balloelektriske Kurver, »Ballogrammer«, kunde give interessante Oplysninger.

[1] C. Christiansen. Experimentalundersøgelser over Gnidningselektricitetens Oprindelse IV. Det kgl. Danske Videnskabernes Selskabs Forhandlinger 1912. No. 5.

[2] W. Pauli u. Handovski. Biochem. Zeitschr. 18. 340. 1909.

[3] Johanne Christiansen. Biochem. Zeitschr. 47. 226. 1912.

Serumalbumin.

Den første Undersøgelsesrække foretog vi med dialyseret Hesteserum, og det viste sig strax, at det balloelektriske Udslag forandrede sig i meget høj Grad med Concentrationen af den tilsatte Syre eller Base, men ganske vist paa en helt anden Maade, end vi paa Forhaand havde ventet. Den følgende Tabel 1 viser en Sammenstilling af det balloelektriske Udslag med Viscositet og Reaktion af samme Opløsning. Der undersøgtes fire Æggehvideconcentrationer, nemlig 0,48%, 0,8%, 1,6% og 2,4% Serumalbumin (Concentrationen bestemt ved Multiplication af Kvælstofprocenten med 6,25). Viscositeten bestemtes kun for Concentrationen 1,6% og angives i Secunder. Den bestemtes ved 20° i Vandbad med et Viscosimeter, der for Vand gav Værdien 32,5 Sek. Reaktionen bestemtes kun groft ved Hjælp af Kongopapir (*K*). Positiv Reaktion paa dette betyder, at Brintionexponenten er $< 4{,}6$, negativ, at den er $> 4{,}6$[1]. Om der var fri Saltsyre tilstede eller ej, bestemtes ved Hjælp af Günzburgs Reagens (*G*). Balloelektriciteten bestemtes paa den af C. Christiansen[2] tidligere angivne Metode. Til Sammenligning benyttedes altid en Kontrol, bestaaende af lige Dele 2 mol. Alkohol og $^n/_{10}$ *KCl*, saaledes at Udslaget af denne Vædske altid sattes lig 50. I 1. Colonne er for hver Albuminconcentration angivet Opløsningens Concentration af Saltsyre.

Tabel 1 og den tilhørende Fig. 1 viser, at Balloelektriciteten, der i Begyndelsen er positiv, ved Tilsætning af ganske smaa Mængder Saltsyre bliver først aftagende indtil 0, derefter negativ, naar et Minimum, for derefter atter at stige jævnt indtil positive Værdier.

[1] Johanne Christiansen. Bestemmelse af fri og bunden Saltsyre i Maveindhold 1912.

[2] l. c.

Tabel 1. Serumalbumin.

0.48 % Serumalb.			0.8 % Serumalb.				1.6 % Serumalb.				2.4 % Serumalb.		
$10^2 \cdot C_{HCl}$	Ballo-elektr.	Reaktion	$10^2 \cdot C_{HCl}$	$10^2 \cdot C_{NaCl}$	Ballo-elektr.	Reaktion	$10^2 \cdot C_{HCl}$	Ballo-elektr.	Viscositet	Reaktion	$10^2 \cdot C_{HCl}$	Ballo-elektr.	Reaktion
0	8	..	0	0	8,7	..	0	10	35,4	..	0	11,2	..
0,05	4,2	$\div K$	0,2	0	$\div$ 4,5	$\div K$	0,2	$\div$ 0,2	..	..	0,4	$\div$ 1,3	..
0,1	$\div$ 4,9	$+ K$	0,25	0	$\div$ 6,2	$+ K$	0,3	$\div$ 3,3	..	..	0,5	$\div$ 1,8	..
0,15	$\div$ 6	..	0,3	0	$\div$ 5,1	..	0,4	$\div$ 4	..	$\div K$	0,6	$\div$ 2,4	$\div K$
0,2	$\div$ 5,8	..	0,5	0	0,2	..	0,5	$\div$ 4,5	..	$+ K$	0,7	$\div$ 2,2	$+ K$
0,3	$\div$ 2,2	..	1	0	10,9	..	0,6	$\div$ 3,3	..	..	0,8	$\div$ 1,3	..
..	..	..	0	5	2,4	..	0,8	1,8	..	..	0,9	$\div$ 0,6	..
..	..	..	0,25	5	5,8	..	1	6,7	39,5	..	1	$+$ 1,1	..
..	..	..	0,5	5	14,3	..	1,5	10,8	..	..	..	..	..
..	..	..	..	..	..	..	2	17,6	49,5	$\div g$	..	..	..
..	..	..	..	..	..	..	2,2	20,9	..	$+ g$	..	..	..
..	..	..	..	..	..	..	2,4	23,6	..	..	..	..	..
..	..	..	..	..	..	..	2,6	25	..	..	..	..	..
..	..	..	..	..	..	..	2,8	27	..	..	..	..	..
..	..	..	..	..	..	..	3	28,4	47,9	..	..	..	..
..	..	..	..	..	..	..	4	31	..	..	..	..	..
..	..	..	..	..	..	..	5	32	45	..	..	..	..

Viscositeten viser en jævnt stigende Kurve, der naar Maximum i det Punkt, hvor der optræder fri Syre, hvor altsaa netop al Æggehviden har forbundet sig med Saltsyre. Da Saltsyreconcentrationen ved dette Maximum er 0,02 n,

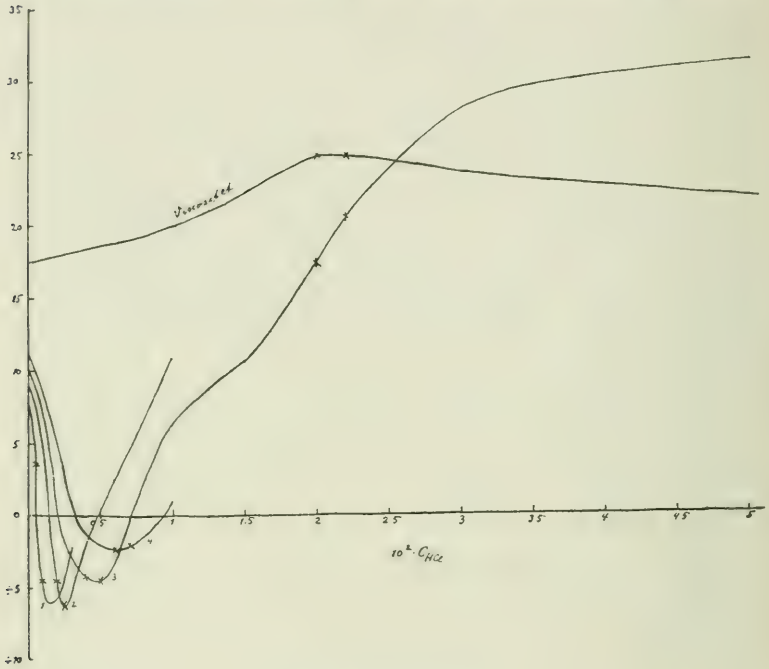


Fig. 1. Serumalbumin.

maa dette altsaa ogsaa betragtes som den omtrentlige Normalitet af 1,6 % Serumalbumin, hvis Molekyle altsaa bliver et Multiplum af 16 : 0,02 = 800.

Svarende til dette Viscositetsmaximum ses ingen Discontinuitet i den tilsvarende balloelektriske Kurve (for 1,6 % Serumalbumin, Kurve 3 paa Figuren), kun bliver Stigningen i Nærheden af dette Punkt noget mindre brat.

Det interessante ved Kurverne knytter sig altsaa ikke, som ventet, til den Saltsyreconcentration, hvor Serum-

albumin er mættet med Saltsyre, men derimod til det højst ejendommelige første Forløb af Ballogrammerne.

Det Synspunkt, der naturligvis strax paatrængte sig, var Muligheden af, at det isoelektriske Punkt af Serumalbumin var Aarsagen til den stærke Svingning fra positiv til negativ Balloelektricitet. Til at bedømme dette var Reaktionen paa Kongopapir en ganske god Vejledning. Omslaget hermed ligger nemlig ved $p_H = 4{,}6$, mens det isoelektriske Punkt for Serumalbumin efter Michaëlis og Davidsohn[1] ligger omtrent ved $p_H = 4{,}7$.

Betragter man de fire Kurver paa Fig. 1, hvor Grænserne for Kongoomslaget er angivet ved Kryds, under dette Synspunkt, ser man, at Kongoomslaget for de stærkeste Concentrationer (1,6 og 2,4%, henholdsvis Kurve 3 og 4) ligger i Nærheden af Minimum, medens det ved de to laveste Concentrationer Kurve 1 og 2 ligger noget til venstre for dette. Da det isoelektriske Punkt som nævnt ligger ved en noget lavere Brintionconcentration end Kongoomslaget, kan det efter disse Forsøg ikke falde sammen med Minimum af Ballogrammerne. Derimod kunde det tænkes at falde sammen med Kurvernes første Skæringspunkt med Abscisseaxen, der jo er et balloelektrisk Neutralpunkt, og dette vilde jo ogsaa være teoretisk bedre forstaaeligt.

Ved de senere Forsøg med Ægalbumin og Aminosyrer rettedes derfor Opmærksomheden særlig mod dette Punkt.

Før vi omtaler Forsøg med andre Stoffer, skal imidlertid for Fuldstændigheds Skyld anføres spredte Forsøg af anden Art, der foretoges med det samme Serumalbumin.

Ved Tilsætning af Natron i voxende Concentration til 1,6% Serumalbumin steg Balloelektriciteten jævnt fra 10 til 12. Den højeste undersøgte Natronconcentration var 0,05 n.

[1] Biochem. Zeitschr. 33. 456. 1911.

Ved Tilsætning af Natriumchlorid til 1,6 % Serumalbumin sank Balloelektriciteten meget hurtig ned til 3, hvor den holdt sig konstant ved Concentrationer mellem 0,005 og 0,05 *n NaCl*.

I Forsøgene med 0,8 % Serumalbumin tilsattes i nogle Tilfælde, som det ses af Tabel 1, foruden Saltsyre ogsaa Natriumchlorid. Det viste sig, at Natriumchlorid, der som ovenfor omtalt formindskede Udslaget af det neutrale Serum, i betydelig Grad forøger Udslaget af det sure Serum.

I andre Forsøg (Tabel 2) tilsattes til 1,6 % Serumalbumin Svovlsyre eller Eddikesyre i Stedet for Saltsyre.

Tabel 2. 1.6 % Serumalbumin.

$10^2 \cdot C_{H_2SO_4}$	0	0,2	0,4	0,6	0,8	1	1,5	2	2,5	3	3,5	4	5
Balloelektr.	9,5	3,8	÷ 0,5	÷ 0,8	÷ 0,3	2,8	8,4	14,2	18	22	24	25	27,4
Reaktion	..	÷ *K*	+ *K*	..	..	..	..	÷ *g*	÷ *g*	..	..	..	..

$C_{C_2H_4O_2}$	0	0,02	0,05	0,2	0,4	0,62	1,25	2,5	..	..	..	..	..
Balloelektr.	6	15,6	÷ 6,4	÷ 6	1,2	30	38	37	..	..	..	..	..
Reaktion	÷ *K*	+ *K*	..	..	..	..	..	..	..	..	..	..	..

Svovlsyre giver lignende Resultater som Saltsyre, men Kurven forløber noget fladere. Kongo-Omslaget ligger meget nær ved det første Skæringspunkt med Nullinien.

Forsøgene med Eddikesyre gav ejendommelige Resultater, idet Balloelektriciteten ved Tilsætning af smaa Mængder Eddikesyre først stiger til et Maximum, derefter falder til 0, saa bliver negativ og tilsidst atter stiger. Kongo-Omslaget ligger nærmest ved Maximum. Denne Kurve er utvivlsomt delvist betinget af, at Eddikesyren selv er et stærkt hyperballisk Stof[1], og dens Kurve er derfor endnu vanskeligere at tyde end Ballogrammerne med de aballiske Syrer.

[1] C. Christiansen l. c.

Witte-Pepton.

Det næste Stof, der undersøgtes, var Witte Pepton (Tabel 3). Dette Stof gav i sig selv et ualmindelig stort balloelektrisk Udslag, nemlig i 2% Opløsning 57, mens 2% Serumalbumin (interpoleret af Værdierne i Tabel 1) gav Balloelektriciteten 10,6 og 2% Ægalbumin (extrapoleret af senere omtalte Forsøg) gav et Udslag paa ca. 7.

Tabel 3. 2% Witte-Pepton.

$10^2 \cdot C_{HCl}$	0	1	1.5	2	3	3,5	4	4,5	5	..
Balloelektr.	57	69	73	76	77	85	86	91	94	..
$10^2 \cdot C_{NaOlt}$	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
Balloelektr.	63	72	77	82	85	86	91	93	95	97
$10^2 \cdot C_{NaCl}$	1	2,5	..	..	..	..	..	..	..	..
Balloelektr.	78	92	..	..	..	..	..	..	..	..

Tabellen viser, at Tilsætning af Saltsyre og Natron til 2% Witte Tepton giver jævnt stigende balloelektriske Værdier, og noget Minimum findes ikke. De faa Forsøg med Natriumchlorid, der findes i Tabellen, viser, at ogsaa Salt forøger Witte Peptonets Balloelektricitet og det i endnu højere Grad end Saltsyre og Natron.

Witte-Pepton undersøgtes kun i denne ene Concentration (2%), og det er derfor — i Analogi med de følgende Æggehvideundersøgelser — ikke udelukket, at et Minimum maaske ved stærkere Fortynding vilde være paaviseligt. Yderligere Forsøg med dette Stof var dog paa Grund af dets Urenhed ikke af stor Interesse.

Æg-Albumin.

Det næste Stof, der toges i Arbejde, var det af Professor S. P. L. Sørensen fremstillede rene Æggehvide-Albumin, som han velvilligt overlod os en Portion af. Det var nemlig

Tabel 4. S. P. L. Sørensens Æggehvide

0,1—25 cc			0,4—25 cc			0,8—25 cc		
0,0179 % Ægalb.			0,0715 % Ægalb.			0,143 % Ægalb.		
$10^3 \cdot C_{H_2SO_4}$	Ball.	p_H	$10^3 \cdot C_{H_2SO_4}$	Ball.	p_H	$10^3 \cdot C_{H_2SO_4}$	Ball.	p_H
0	7,6	..	0	6,25	5,16	0	5,8	5,1
0,08	1,8	..	0,08	2,6	4,77	0,08	0,9	4,85
0,16	÷ 4,1	..	0,16	÷ 6	..	0,16	÷ 4,7	4,7
0,24	÷ 4,3	..	0,24	÷ 6,3	..	0,32	÷ 5,1	..
0,32	÷ 3,4	..	0,32	÷ 6,1	..	0,48	÷ 4,6	..
0,48	÷ 2,3	..	0,48	÷ 4,5	..	0,64	÷ 3,8	..
..	..	..	..	..	..	0,8	÷ 3,5	..
..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..

med Henblik paa Muligheden af balloelektrisk at kunne paavise det isoelektriske Punkt af stor Interesse at arbejde med et Stof, der var ganske frit for Urenheder, der eventuelt kunde mætte den tilsatte Syre. Det isoelektriske Punkt er jo blandt andet ejendommeligt derved, at det for forskellige Concentrationer af Ampholyten kræver ganske samme Syremængde for at naa den for Punktet karakteristiske Brintionconcentration. De herpaa rettede Forsøg havde for Serumalbumin givet utilfredsstillende Resultater, idet der, som Fig. 1 viser, for de fire anvendte Fortyndinger maatte anvendes forskellige Syremængder for at naa begge de Punkter (det negativt balliske Minimum og det balloelektrisk neutrale Skæringspunkt), der kunde være Tale om at udpege som isoelektrisk Punkt, og da dette muligvis kunde skyldes Urenheder, foretoges de følgende Forsøg med det rene Æg-Albumin.

Ifølge Meddelelse fra Carlsberg Laboratoriet inde-

$0.5113\,nN$ = 4,48 % Æggehvide (Faktor 6,25).

2—25 cc			4—25 cc			10—25 cc		
0,359 % Ægalb.			0,718 % Ægalb.			1,79 % Ægalb.		
$10^3 \cdot C_{H_2SO_4}$	Ball.	p_H	$10^3 \cdot C_{H_2SO_4}$	Ball.	p_H	$10^3 \cdot C_{H_2SO_4}$	Ball.	p_H
0	4,7	4,92	0	4,5	4,83	0	6	4,82
0,25	÷ 1,8	4,69	0,5	1,2	4,55	0,5	5,4	4,7
0,5	÷ 2,8	..	2	0,6	..	2	4,7	4,42
2	÷ 1,4	..	4	1,3	..	4	4,7	..
4	0,6	..	8	2,6	..	8	4,7	..
16	2,6	..	16	3,9	..	16	5,6	.
$10^3 \cdot C_{NH_3}$	..	..	32	4,8	..	32	7,6	
0,5	8,3	..	$10^3 \cdot C_{NH_3}$	..	..	$10^3 \cdot C_{NH_3}$	..	..
2	9,6	..	0,5	7	5,09	0,5	6,2	4,92
4	9,6	..	2	9,8	..	2	7,6	..
..	..	..	4	10,1	..	..	..	..

holdt dette 4,48 % Æggehvide (Concentrationen beregnet efter Kvælstofmængden, multipliceret med 6,25) og som eneste Urenhed $11{,}64 \cdot 10^{-3}$ mol. $(NH_4)_2\,SO_4$. 6 Concentratrioner undersøgtes balloelektrisk, nemlig 0,0179 %, 0,0715 %, 0,143 %, 0,359 %, 0,718 % og 1,79 %; disse er paa Figurerne mærket 1—6. Som Syre benyttedes Svovlsyre. Resultaterne findes i Tabel 4 og Fig. 2 –3, hvor Ordinaten betyder Balloelektriciteten, Abscissen Svovlsyreconcentrationen. Enheden for Abscissen er i Fig. 3 50 Gange mindre end i Fig. 2 for at tillade Tegning af Kurverne 4—6 i fuld Udstrækning.

Figurerne viser, at alle 6 Kurver har et tydeligt Minimum, men kun for de fire mindst concentrerede Opløsninger viser dette Minimum balloelektrisk negative Værdier. Den lavest liggende Kurve er No. 2; Kurve 1 ligger igen højere end baade No. 2 og 3.

Det interessante Spørgsmaal var, om det første Skærings-

punkt med Abscisseaxen, der jo ved de foreløbige Forsøg med Serumalbumin var udpeget som det mulige isoelektriske Punkt, laa ved samme Syreconcentration for de forskellige Æggehvideconcentrationer, saaledes som det teoretisk kræves af det isoelektriske Punkt. At dette meget nær er Tilfældet,

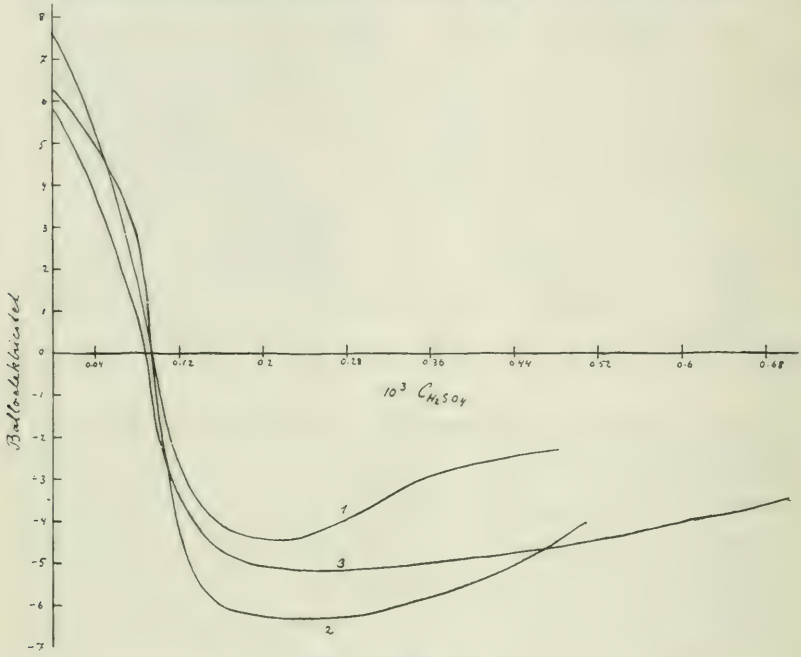


Fig. 2. Ægalbumin.

ses af Kurverne 1—3 paa Fig. 2; skønt Æggehvideconcentrationen her varierer tidobbelt (fra 0,0179—0,143%), saa ligger de omtalte Skæringspunkter meget nær ved hinanden, nemlig ved $10{,}4 \cdot 10^{-5}$, $10{,}4 \cdot 10^{-5}$ og ved $9{,}2 \cdot 10^{-5}$ *n* H_2SO_4. Hvad de øvrige 3 Kurver angaar (No. 4—6), da ligger Skæringspunktet for Kurve 4 et eller andet Sted mellem 0 og $25 \cdot 10^{-5}$, men kan ikke angives nærmere paa Grund af manglende Bestemmelser. Kurve 5 og 6 skærer overhovedet ikke Abscisseaxen; dette er et Forhold, der ved

balloelektriske Forsøg ofte gentager sig. Kun ved ret stærke Fortyndinger kommer de karakteristiske Kurver frem, muligvis fordi Indflydelsen af tilstedeværende Salte eller andre Urenheder herved elimineres.

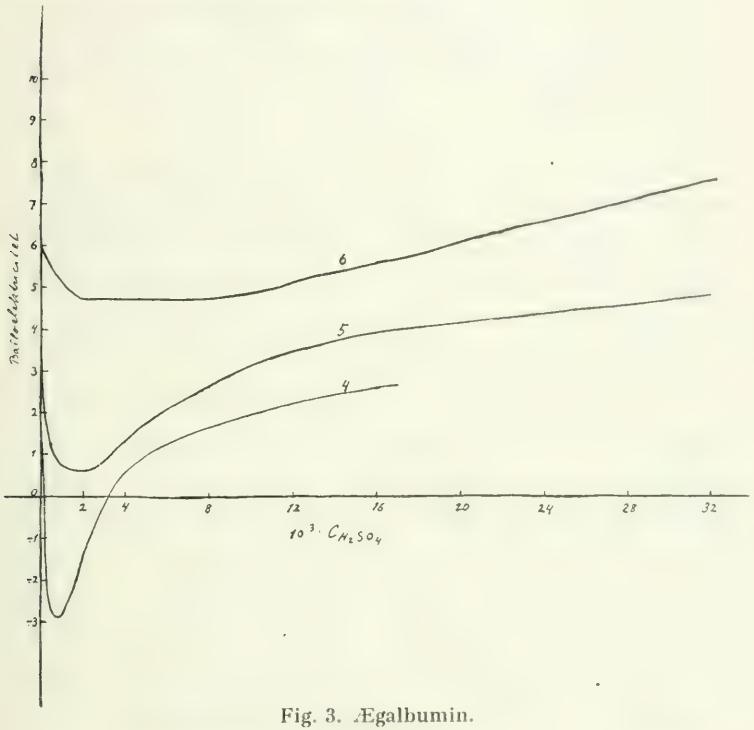


Fig. 3. Ægalbumin.

Det andet karakteristiske Kendetegn for det isoelektriske Punkt er, at det for forskellige Æggehvideconcentrationer ligger ved samme Brintionconcentrion. Undersøgelsen af Brintionconcentrationen af disse Opløsninger har vi ikke selv foretaget, men paa Grundlag af de paa Carlsberg Laboratoriet udførte Brintionconcentrationsmaalinger af Blandinger af ammoniumsulfatholdige Ægalbuminopløsninger og lidt Svovlsyre eller Ammoniak kunde Brintion-

concentrationen grafisk, om end ikke bestemmes nøjagtigt, saa dog skønnes i en Del af de benyttede Opløsninger. De paa denne Maade tilnærmelsesvis fundne Værdier for p_H findes angivne i Tabel 4.

Det ses heraf, at det første Skæringspunkt med Abscisseaxen for 0,0715 % Æggehvide ligger i Nærheden af $p_H = 4{,}77$, for 0,143 % Æggehvide mellem 4,7 og 4,85 og for 0,359 % Æggehvide mellem 4,69 og 4,92.

Da S. P. L. Sørensen[1] ad anden Vej har fundet, at det isoelektriske Punkt for det rene Ægalbumin ligger ved $p_H = 4{,}803$, synes det altsaa, at det første Skæringspunkt af den balloelektriske Kurve med Abscisseaxen opfylder ogsaa den anden Hovedbetingelse, som stilles til det isoelektriske Punkt.

Muligheden af, at det isoelektriske Punkt af amphotere Stoffer under visse Betingelser (stærk Fortynding af Opløsningen) var balloelektrisk neutralt, altsaa under Sprøjtningen udviklede lige mange positive og negative Elektricitetsbærere, var med Henblik paa den endnu dunkle Teori for Balloelektriciteten saa interessant, at vi ansaa det for paakrævet at undersøge ogsaa Aminosyrerne i denne Retning. Det isoelektriske Punkts Beliggenhed er velkendt for mange af disse, og ved samtidige Brintionconcentrationsbestemmelser og balloelektriske Bestemmelser var det derfor sandsynligt, at man kunde opnaa interessante Resultater.

Der optraadte imidlertid under dette Arbejde store tekniske Vanskeligheder, da Opløsninger af Aminosyrer viste sig stærkt foranderlige i balloelektrisk Henseende ved mindste Spor af Urenheder. Balloelektriciteten af en Opløsning af n Glycin stiger saaledes ved Berøring med Filtrerpapir fra 5,5 til 12,5. Forholdet var altsaa ganske som det tidligere

[1]) Meddelelser fra Carlsberg Laboratoriet. Bd. 12. 1915—17.

er paavist for Vandets Vedkommende af C. Christiansen[1]. Under de samme Forhold steg nemlig Balloelektriciteten af Ledningsevnevand fra 11,5 til 20; en Opløsning af $^{n}/_{10}$ NaCl gav derimod baade før og efter Berøring med Filtrerpapir ganske samme Balloelektricitet ($= 0{,}5$).

Endvidere maatte man tage Hensyn til følgende Mulig heder for Fejlkilder: 1) Luftens Kulsyre 2) Glasalkalier 3) det til Fortyndingerne anvendte Vand.

Det viste sig nemlig at spille en meget stor Rolle, om Opløsningerne var fremstillet med alm. destilleret Vand, med kulsyrefrit destilleret Vand eller med dobbelt destilleret Ledningsevnevand, og det viste sig, at saavel Balloelektricitet som Brintionconcentration forandrede sig ved Henstand i almindelige Reagensglas, medens dette ikke var Tilfældet, naar der anvendtes Reagensglas af Jenaglas.

Efter en stor Mængde indledende Forsøg, der maatte kasseres, foretoges derfor Opløsningen af de rene Stoffer i Vand, der var destilleret først med Kaliumpermanganat og Svovlsyre gennem Svalerør af Jenaglas, derefter med Barytvand gennem Svalerør af Sølv og tilsidst udkogt og kulsyrefrit opbevaret i damprensede Jenakolber med tilhørende Buretter af Jenaglas. Ledningsevnen af dette Vand var $1 \cdot 10^{-6}$.

Alle Maalekolber, Pipetter osv., der benyttedes til Fremstillingen og Opbevaringen af Opløsningerne var af Jenaglas. Til de balloelektriske Forsøg konstruerede C. Christiansen et særligt lille Ballometer, fuldstændig af Jenaglas og Platin og indrettet saaledes, at den samme Vædskemængde under Undersøgelsen stadig cirkulerede i Apparatet og bagefter kunde benyttes til Bestemmelse af

[1] C. Christiansen: Experimentalundersøgelser over Gnidningselektricitetens Oprindelse III. Det kgl. Danske Videnskabernes Selskabs Forhandlinger. 1911. No. 3.

Brintionconcentrationen. Ved de sædvanlige Ballometre gaar nemlig Vædsken til Spilde, men af Hensyn til de nævnte Fejlkilder var det ved de planlagte Undersøgelser af stor Vigtighed, at baade Balloelektriciteten og Brintionconcentrationen kunde bestemmes paa samme Vædske.

Apparatet, som ses paa Fig. 4, var indrettet paa følgende Maade: *A* er Sprøjteroret, der er indsmeltet i den større Beholder, *B*, der indeholder Platinpladen, *P*, som er i ledende Forbindelse med Kvadrantelektrometret. *S* er Skorstenen, hvorigennem de fraspaltede Elektricitetsbærere undviger. Gennem Røret, *b*, opsuges den Vædske, der skal undersøges, af den gennem *a* passerende Luftstrøm, og efter at have sprøjtet mod Platinpladen dryppes den atter tilbage gennem Røret *C*. Reagensglasset, *R* er af Jenaglas og indeholder 10 cc. af den Opløsning, der skal undersøges. Umiddelbart efter den balloelektriske Undersøgelse foretages kolorimetrisk Brintionconcentrationsbestemmelse ad modum S. P. L. Sørensen[1].

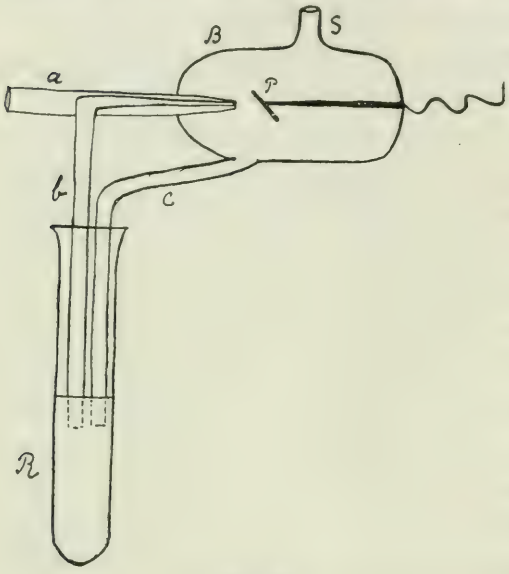


Fig. 4.

Elektrometrisk Brintionconcentrationsbestemmelse var umulig paa Grund af Vædskernes minimale Ledningsevne,

[1] Meddelelser fra Carlsberg Laboratoriet. 1909. Bd. 8.

og det var heller ikke tilraadeligt at tilsætte Stødpuder eller Salte, da disse forandrer Vædskens balloelektriske Forhold meget stærkt. Da Vædskerne kun bestod af en Opløsning af en Ampholyt, tilsat en ringe Mængde Syre eller Base, er de vundne Tal for Brintionconcentrationerne næppe saa sikre, som det var at ønske.

Glycin $CH_2\ (NH_2) \cdot COOH$.

Den første Aminosyre, der undersøgtes, var Glycinet, hvis isoelektriske Punkt ligger ved $p_H = 6{,}1$[1], og der mødte os strax den Skuffelse, at Ballogrammet var af en helt anden Karakter end de tidligere omtalte Ballogrammer for Æggehvidestoffer. Det viste ingen bratte Svingninger fra positiv til negativ Balloelektricitet, og stigende Syretilsætning bevirkede kun en jævn Synken af den balloelektriske Kurve henimod Nullinien. Forsøgene gentoges talrige Gange med de følgende Fortyndinger af Glycinet: $^n/_{1600}$, $^n/_{400}$, $^n/_{100}$, $^n/_{20}$ og $^n/_2$ og med forskellige Syrer (Saltsyre, Eddikesyre, Phosphorsyre, Borsyre) uden at der opdagedes nogensomhelst Uregelmæssighed i Kurverne.

Tabel 5 indeholder nogle af disse Forsøg med Saltsyre- og Natron-Tilsætning og de tilhørende Brintionexponenter.

Glycinopløsningerne alene uden Tilsætninger viste følgende Balloelektricitet: 2n : 11, n : 12,8, $^n/_2$: 17, $^n/_4$: 18,6, $^n/_{16}$: 12,5, $^n/_{32}$: 26, $^n/_{64}$: 26, $^n/_{128}$: 12,5, $^n/_{1024}$: 10: Ledningsevnevand : 12. Der findes altsaa et Maximum ved omtrent $^n/_{50}$ Glycin.

Det er bekendt fra C. Christiansens tidligere Undersøgelser, at Syrer, Baser og Salte nedsætter Vandets Balloelektricitet stærkt, saaledes at den ved en vis — ret varierende

[1] L. Michaelis. Biochem. Zeitschr. 47. 250. 1912.

- Concentration bliver 0 og ved højere Concentration bliver svagt negativ.

Vand kan jo opfattes som et amphotert Stof, og parallelle Undersøgelser foretoges derfor ogsaa med Glycinet. Vi fandt, at $^n/_{100}$ Glycin gav balloelektrisk Neutralitet med henholdsvis $^n/_{256}$ *HCl*, $^n/_{32}$ *NaOH* og $^n/_{32}$ *NaCl*, og $^n/_2$ Glycin med $^n/_{64}$ *HCl*, $^n/_8$ *NaOH* og $^n/_{32}$ *NaCl*. Det synes altsaa, som om den Syre- og Base-Concentration, der giver balloelektrisk Neutralitet, er noget afhængig af Glycin-Concentrationen, men at Natriumchlorid-Concentrationen er uafhængig af denne.

Indeholder Glycinopløsningen saa meget Salt, at dens Balloelektricitet ligger i Nærheden af Neutralpunktet, har Tilsætning af Syre eller Base ingensomhelst Indflydelse.

Da C. Christiansen tidligere har vist, at Alkoholtilsætning til visse Stoffer har stor balloelektrisk Virkning, prøvedes dette overfor Glycin.

En Tilsætning til $^n/_{10}$ Glycin af lige Dele 2 *n* Alkohol forøgede Balloelektriciteten af $^n/_{20}$ Glycin fra 10,6 til 23,6; ogsaa sure og alkaliske Glycinopløsninger fik en stærk Forøgelse af deres Balloelektricitet ved Tilsætning af Alkohol.

Tabel 5.

0,01 *n* Glycin						0,01 *n* Alanin					
$10^3 \cdot C_{HCl}$	Ball.	p_H	$10^3 \cdot C_{NaOH}$	Ball.	p_H	$10^3 \cdot C_{HCl}$	Ball.	p_H	$10^3 \cdot C_{NaOH}$	Ball.	p_H
0	17,8	6,75	0,1	13,9	7,6	0	23,5	6,53	0,1	21	7,0
0,1	14,9	5,44	1	9,4	8,1	0,05	13,7	5,5	0,2	20	7,55
0,2	10,6	4,30	2	6,7	8,8	0,1	8,4	4,9	0,3	17,5	7,73
0.5	6,1	3,95	5	3,9	9,16	0.15	6,1	4,5	0.4	15,6	8,0
5	0,3	2,9	50	÷0,4	>16	0,2	5,6	4,3	10	0,6	10
50	0	1,53	..	..	..	0,3	4,4	4,05	..	..	..
..	..	..	..	..	..	0,5	3,4	3,95	..	..	..
..	..	..	..	..	..	1	0,7	3,65	..	..	..

Alanin. $CH_3 \cdot CH\ (NH_2) \cdot COOH$.

Heller ikke Alaninets isoelektriske Punkt kunde opdages balloelektrisk. De undersøgtes Alaninconcentrationerne $^n/_{1000}$, $^n/_{500}$, $^n/_{100}$, $^n/_{40}$ og $^n/_{20}$, og baade Saltsyre, Eddikesyre, Phosphorsyre og Borsyre prøvedes, men Resultaterne var omtrent som ved Glycin. Tabel 6 viser en af Forsøgsrækkerne.

Tilsætning af Alkohol til en Concentration af 0,2 *n* Alkohol forøgede Udslaget af baade neutrale, sure og alkaliske Blandinger til over det dobbelte.

Salt nedsætter Alaninets Balloelektricitet.

β-Alanin. $CH_2(NH_2) \cdot CH_2 \cdot COOH$.

Dette Stof undersøgtes kun i Concentrationen $^n/_{100}$ og kun med Phosphorsyre og Natron. Det gav meget smaa Udslag, idet den rene $^n/_{100}$ β-Alanin-Opløsning, der gav det største Udslag, kun havde Balloelektriciteten 3, og dette nedsattes yderligere af saavel Syre som Base.

α-Amino-Smørsyre. $CH_3 \cdot CH_2 \cdot CH\ (NH_2) \cdot COOH$.

Ballogrammet af dette Stof for Concentrationen 0,01 n ses paa Fig. 5. Den Kurve, der er mærket: Alkohol, er forskellig fra den anden derved, at der foruden Aminosyre og Saltsyre ogsaa er 0,02 *n* Alkohol tilstede i alle Blandingerne. Abscissen forestiller p_H, Ordinaten det balloelektriske Udslag.

Figuren viser det højst mærkelige Forhold, at de to Kurver med og uden Alkohol skærer hinanden paa Nullinien ved $p_H = 4{,}6$. Før dette Punkt forøger Alkoholen den positive, efter det forøger den den negative Balloelektricitet. For nærmere at oplyse dette interessante Fænomén, gennemførtes Bestemmelser af samme Art ogsaa i Concen-

trationerne: 0,015, 0,005, og 0,002 n Amino-Smørsyre. Resultaterne er samlede i Tabel 6. Det viste sig, at den mest fortyndede af disse Opløsninger (0,002 n) ikke skar Nullinien ved nogen af de anvendte Syreconcentrationer,

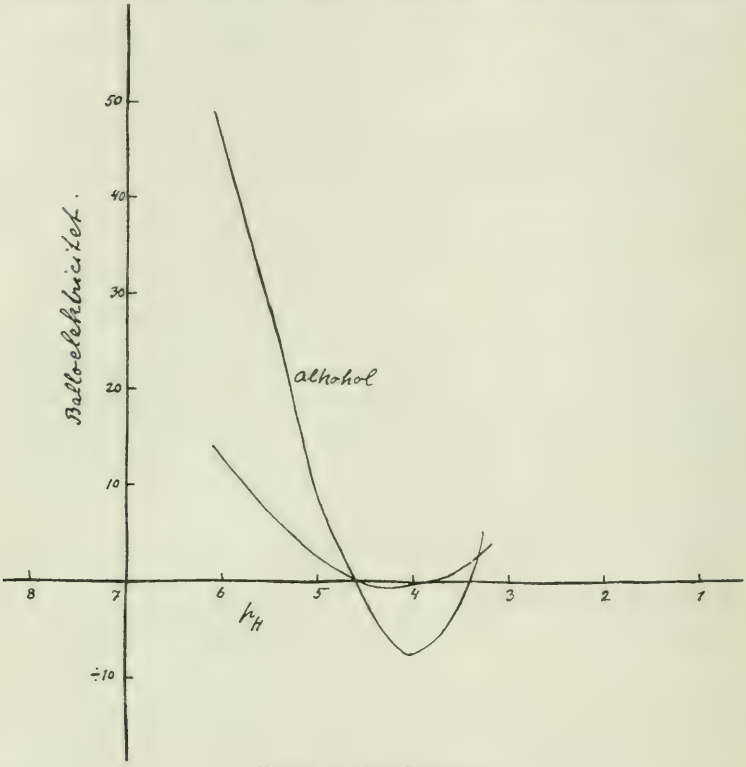


Fig. 5. Amino-Smørsyre.

hvorimod Kurven for 0,015 n Amino-Smørsyre skærer ved $p_H = 5$, dens tilsvarende Alkoholkurve ved 4,9, og Kurven for 0,005 n Amino-Smørsyre berører Nullinien ved $p_H = 4{,}3$, mens dens tilsvarende Alkoholkurve skærer ved 4,15. Disse Punkter, der for Concentrationen 0,01 n Amino-Smørsyre faldt fuldstændig sammen, ligger altsaa ogsaa for disse Concentrationer saa nær hinanden, at Forskellen utvivlsomt

Tabel 6. Amino-Smørsyre.

0,002 *n* A.-S.					0,005 *n* A.-S.					0,01 *n* A.-S.					0,015 *n* A.-S.				
	uden Alkohol		m. 0,2 *n* Alkohol			uden Alkohol		m. 0,2 *n* Alkohol			uden Alkohol		m. 0,2 *n* Alkohol			uden Alkohol		m. 0,2 *n* Alkohol	
$10^4 \cdot C_{HCl}$	Ball.	p_H	Ball.	p_H	$10^4 \cdot C_{HCl}$	Ball.	p_H	Ball.	p_H	$10^4 \cdot C_{HCl}$	Ball.	p_H	Ball.	p_H	$10^4 \cdot C_{HCl}$	Ball.	p_H	Ball.	p_H
0	11,2	6,9	36	6,8	0	13,1	6,12	45	6,12	0	13,7	6,12	49	6,12	0	10	6,24	41	6,24
0,6	11,2	6,2	14,4	5,5	0,5	4,8	5,15	24	5,35	0,5	5	5,31	26	5,44	0,5	3,6	5,5	22	5,5
0,8	4,2	5,1	11,2	5,4	1	1,2	4,6	10,6	4,7	1	1	4,8	9,1	5,03	1	0	5,0	5,3	5
1	3,4	4,6	9,4	4,4	2	0	4,3	0,6	4,3	2	÷ 0,6	4,35	÷ 3,1	4,45	1,5	÷ 2,6	4,8	÷ 4,4	4,8
2	1,6	4,3	5,9	3,9	4	0,5	3,85	÷ 0,9	3,85	4	0	4,05	÷ 7,8	4,05	5	÷ 5	4,3	÷ 16,6	4,3
5	0,9	3,7	4,6	3,7	..	..	..	..	..	10	0,6	3,65	÷ 5	3,68	15	2,8	3,62	÷ 4,4	3,62
..	..	..	..	..	..	..	..	..	..	20	3,8	3,22	5	3,3	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	40	5,3	3	15	3	..	.			

falder indenfor den anvendte Metodes Fejlgrænser. Alkohol synes altsaa ikke at paavirke Amino-Smørsyren i dens første balloelektriske Nulpunkt. Denne Egenskab kunde meget vel tænkes at være karakteristisk for det isoelektriske Punkt, da andre Undersøgelser af C. Christiansen har vist, at modsat ladede Complexer (her positive og negative Amino-Smørsyre-Joner) ofte giver modsat balloelektrisk Udslag ved Tilsætning af Alkohol. Imod Antagelsen af, at dette Skæringspunkt svarer til det isoelektriske Punkt taler selvfølgelig den noget forskellige Beliggenhed ved forskellige Concentrationer af Amino-Smørsyre.

Amino-iso-Smørsyre. $(CH_3)_2 \cdot C\,(NH_2) \cdot COOH.$

Denne undersøgtes i to Concentrationer: nemlig 0,015 og 0,01 normal, med Saltsyre, og med og uden Alkohol. Resultaterne findes i Tabel 7.

Tabel 7. Amino-iso-Smørsyre.

0,01 *n* A.-i.-S.					0,015 *n* A.-i.-S.				
	uden Alkohol		m. 0,2 *n* Alkohol			uden Alkohol		m. 0,2 *n* Alkohol	
$10^4 \cdot C_{HCl}$	Ball.	p_H	Ball.	p_H	$10^4 \cdot C_{HCl}$	Ball.	p_H	Ball.	p_H
0	15	6,12	55	6,12	0	16,9	6,2	52,5	6,2
0,5	3,4	5,31	26	5,4	0,5	8,8	5,31	25	5,31
1	÷ 1,6	4,83	8,4	4,9	1	1,1	4,95	5,6	4,95
2	÷ 2,4	4,5	÷ 3,8	4,5	1,5	÷ 1,9	4,74	÷ 4,4	4,74
4	÷ 0,6	4,1	÷ 7,5	4,1	10	2,5	3,9	÷ 8,8	3,9
10	1,2	3,7	÷ 4,4	3,7	..	..	..	..	..
20	5,3	3,37	4,4	3,4	..	..	..	..	..
40	7,5	3	16	3	..	..	..	..	..

Kurverne er ikke optegnede, da de fuldstændig ligner Kurven for Amino-Smørsyren. Skæringspunkterne med Nul-

linien ligger for 0,015 n Amino-iso-Smørsyre ved $p_H = 4{,}8$ baade med og uden Alkohol, for 0,01 n ved $p_H = 5$ uden, ved 4,7 med Alkohol. Ogsaa her synes Alkoholen altsaa i

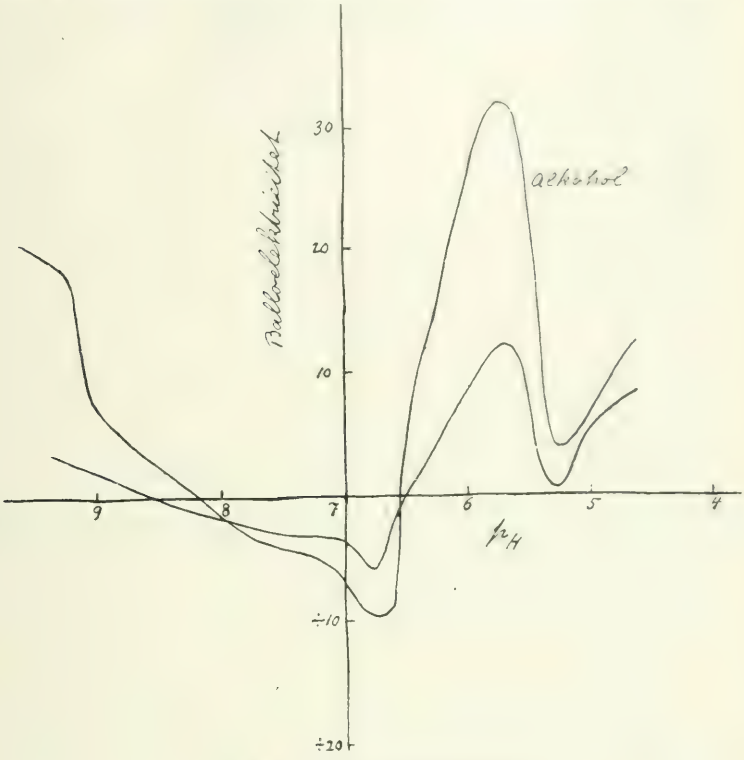


Fig. 6. δ-Amino-Valerianesyre.

Nærheden af det formodede isoelektriske Punkt at være uden Virkning, mens den før og efter dette stærkt foroger henholdsvis det positive og negative balloelektriske Udslag.

δ-Amino-Valerianesyre. $CH_2\,(NH_2)\,CH_2\,CH_2\,CH_2\cdot COOH$.

Denne Aminosyre undersøgtes kun i Concentrationen $^{n}/_{100}$, men baade med Saltsyre og Natron og med Alkoholtilsætning (Tabel 8). Ballogrammet er højst ejendommeligt, idet Stoffet selv i Modsætning til alle de andre Stoffer gav

negativt balloelektrisk Udslag og først ved Tilsætning af Syre eller Base blev positiv balloelektrisk. Ballogrammet skærer derfor, som Fig. 6 viser, Nullinien baade paa den sure ($p_H = 6{,}5$) og den alkaliske Side ($p_H = 8{,}3$). I Nær-

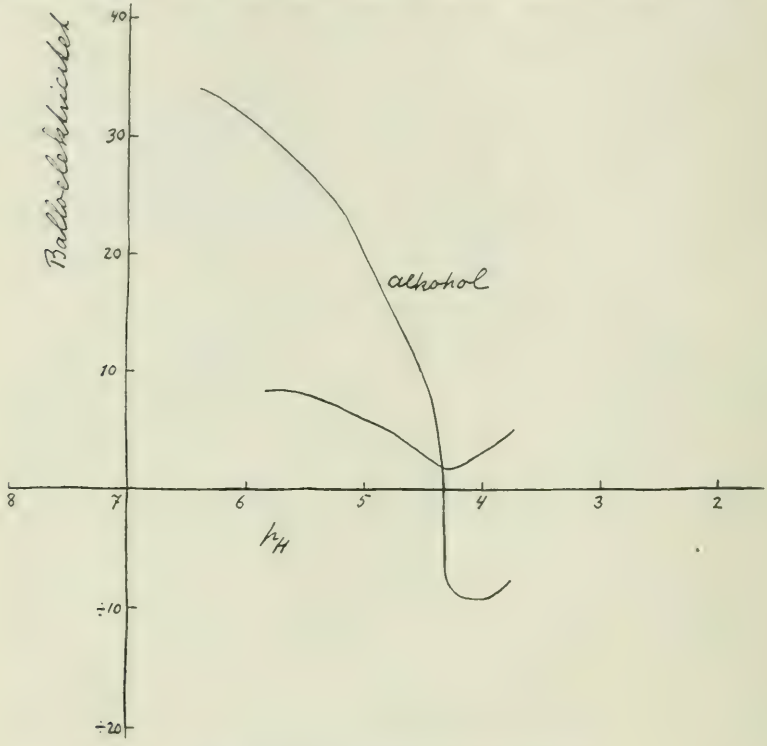


Fig. 7. 0,015 *n* Amino-iso-Valerianesyre.

heden af disse balloelektriske Neutralpunkter har Alkohol ingen eller meget ringe Indflydelse paa Opløsningernes Balloelektricitet, mens det ved andre Concentrationer virker meget stærkt forøgende paa henholdsvis den positive og negative Balloelektricitet.

I to Forsøg tilsattes $^n/_{10}$ Natriumchlorid til sure Opløsninger af Amino-Valerinesyre ved $p_H = 5$ og 4,4. Udslaget nedsattes herved henholdsvis fra 4,4 til 1,3 og fra 8,5 til 1,6.

Tabel 8. 0,01 *n* δ-Amino-Valerianesyre.

	uden Alkohol		m. 0,2 *n* Alkohol			uden Alkohol		m. 0.2 *n* Alkohol	
$10^4 \cdot C_{HCl}$	Ball.	p_H	Ball.	p_H	$10^4 \cdot C_{HCl}$	Ball.	p_H	Ball.	p_H
0	÷ 2,2	6,65	÷ 8,8	6,65	0	÷ 5,9	6,73	÷ 8.7	6,9
0,5	0,6	6,47	5,1	6,47	0,5	÷ 3.4	6,98	÷ 5.9	7,06
1	4,4	6,24	15,6	6,24	1	÷ 3	7,55	÷ 3.4	7.73
2	8,8	6,00	26	6,00	2	0.3	8,6	2.9	8,6
4	12	5,74	32	5.74	4	2.9	9,09	7	9.09
10	0,7	5,25	3,9	5,25	10	3.4	9.09	12.2	
20	4,4	5,05	5,6	5,1	20	3,3	9,3	18	9,3
40	8,45	4,65	12,5	4,65	40	3,1	9,63	20	9.63

α-**Amino-iso-Valerianesyre** $(CH_3)_2 \cdot CH \cdot CH \cdot (NH_2) \cdot COOH$ undersøgtes i to Concentrationer, nemlig 0,015 *n* og 0,01 *n* med Saltsyre og med og uden Alkohol. De alkoholfri Ballogrammer (se Fig. 7) skærer ikke Nullinien; er der 0,2n Alkohol tilstede, ligger der derimod et Skæringspunkt ved $p_H = 4{,}4$ for 0,015 *n* og ved 4,2 for 0,01 *n* Amino-iso-Valerianesyre. Bestemmelserne findes i Tabel 9.

Tabel 9. α-Amino-iso-Valerianesyre.

0,01 *n* Am.-iso-Valerianesyre					0,015 *n* Am.-iso-Valerianesyre				
	uden Alkohol		m. 0,2 *n* Alkohol			uden Alkohol		m. 0,2 *n* Alkohol	
$10^4 \cdot C_{HCl}$	Ball.	p_H	Ball.	p_H	$10^4 \cdot C_{HCl}$	Ball.	p_H	Ball.	p_H
0	9,4	5,82	31,6	5,85	0	8,4	5,82	34	6,4
0,5	6,1	5,1	22,5	5,2	0,5	6,9	5,2	24	5,2
1	3,7	4,8	14,7	4,74	1	4,8	4,74	17,5	4,89
2	1,3	4,3	1,9	4,25	1,5	3,7	4,55	11,2	4,55
4	1,3	4,15	÷ 6,9	4,15	5	1,7	4,3	÷ 7,8	4,3
7	1,6	3,87	÷ 10,6	3,87	7	3	4,05	÷ 9	4,05
10	4,1	3,62	÷ 9,1	3,62	10	5,1	3,74	÷ 7,8	3,8

Leucin. $(CH_3)_2 \cdot CH \cdot CH_2 \cdot CH\,(NH_2) \cdot COOH.$

Leucin gjordes til Genstand for talrige Undersøgelser, da det viste en meget karakteristisk balloelektrisk Kurve, og da det kunde skaffes til Veje i saa stor Mængde, at det

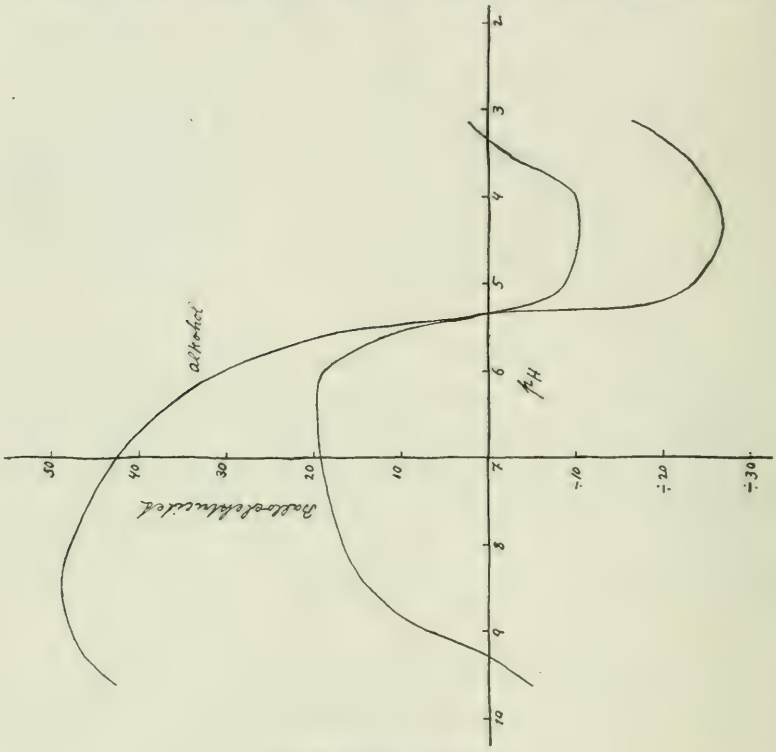


Fig. 8. 0,002 *n* Leucin.

kunde renses omhyggeligt ved Omkrystallisationer, kontrollerede ved Smeltepunktsbestemmelser.

Paa Fig. 8 med tilhørende Tabel 10 ses dets karakteristiske Ballogram med Saltsyre og Natron med og uden Alkoholtilsætning. Leucinconcentrationen er her 0,002 *n*. Det ses, at det første Skæringspunkt med Nullinien (iso-isoelektrisk Punkt?) ligger ved $p_H = 5{,}3$ og $C_{HCl} = 10^{-4} . 0{,}37$.

Tabel 10. 0,002 *n* Leucin.

	uden Alkohol		m. 0,2 *n* Alkohol			uden Alkohol		m. 0,2 *n* Alkohol	
$10^4 \cdot C_{HCl}$	Ball.	p_H	Ball.	p_H	$10^4 \cdot C_{NaOH}$	Ball.	p_H	Ball.	p_H
0	19,4	6,2	41,5	6,2	0,5	17	7,45	43	7,45
0,2	7,5	5,5	23	5,5	1	16,6	7,9	47,5	7,9
0,3	0,4	5,4	7,8	5,5	2	14,4	8,4	48	8,4
0,4	÷ 1,6	5,2	÷ 22	5,2	3	12,8	8,75	49	8,55
0,6	÷ 7,5	5,15	÷ 19	5,2	4	10	8,85	48	8,8
0,8	÷ 9,7	4,5	÷ 23,4	4,55	10	2,2	9,16	47	9,16
1	÷ 10,3	4,45	÷ 22,8	4,45	20	÷ 4,7	9,63	43	9,6
1,5	÷ 10,3	4	÷ 28,5	4	50	÷ 7,8	> 10	33,5	> 10
6	÷ 2,5	3,53	÷ 24	3,53	100	÷ 5	..	27	..
8	÷ 0,3	3,37	÷ 21	3,37	500	÷ 1,5	..	31,2	..
10	1,6	3,22	÷ 17	3,22	..	..	..	.	..

Forsøgsrækker med andre Leucinconcentrationer og andre Syrer gav de følgende Resultater med Hensyn til Beliggenheden af det hypotetiske isoelektriske Punkt:

Tabel 11.

		Skæringspunktets	
C_{Leucin}	Syrens Art	p_H	$C_{\text{Syre}} \cdot 10 \div 5$
0,04 *n*	Saltsyre	5,05	27
0,02	—	5,05	11
0,015	—	5,15	9
0,01	—	5	7,3
0,01	—	5,1	7,2
0,002	—	5,1	7
0,002	—	5,3	3,7
0,01	Eddikesyre	5,05	10
0,002	—	4,9	9
0,01	Phosphorsyre	5,13	8,3
0,002	Trichloreddikesyre	4,4	8

Alle Syreconcentrationerne er beregnede molekylært, og det ses af Resultaterne, hvoraf nogle er fremgaaet af Dobbeltforsøg, anstillede til forskellige Tider med samme Concen-

trationer, at Syreconcentrationerne med to Undtagelser ligger ret nær hinanden.

Det omdebatterede Punkts p_H ligger for næsten alle Forsøgsrækker mellem 5 og 5,3; kun Trichloreddikesyren og tildels Eddikesyren viser her et afvigende Forhold, men dette skyldes vistnok, at disse Syrer i sig selv har meget ejendommelige balloelektriske Virkninger[1].

Ogsaa Citronsyre og Vinsyre undersøgtes og gav lignende Resultater, som dog ikke skal meddeles, da Teknikken ved disse Forsøgsrækker ikke var ganske korrekt.

Virkningen af Tilsætning af Borsyre prøvedes ogsaa, men gav fuldstændig afvigende Resultater. Det tilsattes i Concentrationer indtil 0,25 mol. og p_H varierede mellem 6,2 og 5,25, men Udslaget var stadig positivt og omtrent uforandret det samme som af den rene $^n/_{100}$ Leucin.

Forsøgene med Leucin viste, hvor overordentlig stor Betydning Indførelsen af de tidligere nævnte Forsigtighedsregler ved Fremstillingen og Fortyndingen af Aminosyrerne havde. Det var de uregelmæssige Forsøgsresultater under Undersøgelsen af dette Stof, der oprindelig gav os Anledning til at foretage de nævnte Forbedringer i Teknikken og de følgende Tal, der fandtes før disse Forbedringer, viser, naar man sammenligner dem med Tallene paa S. 27, at der i ikke helt rene Opløsninger kræves betydelig større Syremængder for at naa det hypotetiske isoelektriske Punkt.

Tabel 12.

C_{Leucin}	Syrens Art	$C_{Syre} \cdot 10 \div 5$ v. Skæringsp.
0,0156 *n*	Saltsyre	30
0,01 -	Eddikesyre	100
0,01 -	Phosphorsyre	27
0,01 -	Citronsyre	17
0,01 -	Vinsyre	17

[1] C. Christiansen l. c.

Selve Kurveforløbet og Størrelsen af Udslagene synes at være ens for de forskellige Syrer (undtagen Borsyre og Trichloreddikesyre) og kun at afhænge af Brintionconcentrationen.

Naar $^n/_{100}$ Leucin er mættet med henholdsvis Syre og Base ($^n/_{100}$ Saltsyre eller $^n/_{100}$ Natron) giver det Balloelektriciteten resp. + 13,4 og ÷ 13,4. Tilsættes lige Dele 2 mol. Alkohol, bliver baade det sure og det alkaliske Salt stærkt positive, det sidste stærkest. Indeholdt $^n/_{100}$ Leucin $^n/_4$ Saltsyre, var Balloelektriciteten 9,4; indeholdt $^n/_{100}$ Leucin $^n/_4$ Natron, var Balloelektriciteten ÷ 1,8.

Hvad Balloelektriciteten af elektrolytfri Leucinopløsninger angaar, da stiger denne med Leucinconcentrationen, som det ses af den følgende lille Tabel, der ogsaa indeholder Forsøg med Tilsætning af Alkohol.

Tabel 13.

x			x blandet med lige Dele	
			Vand	2 n Alkohol
$^n/_{20}$ Leucin	Vand	Balloelek.	Balloelek.	Balloelek.
0	10	8,8	14	28,2
0,5	9,5	11,6	11	42,5
1	9	13	11	49,4
2	8	18,5	13,4	60,6
4	6	33,4	17,8	74,6
6	4	48,7	23,4	89
8	2	62,5	32,5	92,5
10	0	76,6	42	95

Tabellen viser, at Alkohol foroger Balloelektriciteten af fortyndede Leucinopløsninger til det tre- eller firedobbelte, mens de stærkere concentrerede Opløsninger ikke paavirkes nær saa stærkt.

Der foretoges foruden de nævnte ogsaa en Del Forsøg under Tilsætning dels af Alkohol, dels af Natriumchlorid ved forskellige Brintionconcentrationer. Disse Forsøgsrækker

anstilledes saaledes, at Concentrationen af Leucin og af Alkohol eller Natriumchlorid var konstant i den enkelte Række, mens C_H varieredes ved Tilsætning af Saltsyre eller Natron. Den grafiske Fremstilling af disse Forsøgsrækker viste, at Alkohol gav langt stejlere, Natriumchlorid langt fladere Ballogrammer end de simple Blandinger, mens Grundformen dog var den samme for dem alle. Alle Kurverne havde eet fælles Punkt, nemlig det omtalte Skæringspunkt, hvis Beliggenhed er mellem $p_H = 5$ og 5.3.

Med Alkohol fremstilledes kun den i Fig. 8 grafisk optegnede Forsøgsrække. Alkoholconcentrationen var $0{,}2\,n$ Skæringspunktet ligger ved $p_H = 5{,}3$ og $C_{HCl} = 3{,}8 \cdot 10^{-5}$.

Med Natriumchlorid fremstilledes de følgende Forsøgsrækker, alle med $0{,}002\,n$ Leucin:

1) $0{,}001\,n$ *NaCl* gav en Kurve, hvis højeste Punkt var 7, og som skar Nullinien ved $p_H = 5{,}12$ og $C_{HCl} = 6 \cdot 10^{-5}$.

2) $0{,}002\,n$ *NaCl*: Højeste Punkt: 3,6; Skæringspunkt ved $p_H = 5{,}1$ og $C_{HCl} = 70 \cdot 10^{-5}$.

3) $0{,}004\,n$ *NaCl*: Højeste Punkt: 2,6; Skæringspunkt ved $p_H = 5{,}15$ og $C_{HCl} = 18 \cdot 10^{-5}$.

4) $0{,}01\,n$ *NaCl*: Højeste Punkt: 0,9; Skæringspunkt ved $p_H = 5{,}2$ og $C_{HCl} = 26 \cdot 10^{-5}$.

Leucinets isoelektriske Punkt skal ligge ved $p_H = 6{,}06$[1], medens det her omtalte Punkt ligger ved $p_H =$ c. 5,2. Ikke desto mindre er der meget, der taler for, at det balloelektriske Nulpunkt ved 5,2 virkelig svarer til Leucinets isoelektriske Punkt. Det er nemlig et for Leucinet absolut karakteristisk Punkt, idet det er uforandret det samme, ligegyldig hvad Leucinconcentration, hvilken Syre, og hvilken Salt- eller Alkohol-Concentration man anvender. Endvidere

[1] Z. f. physikal. Chemie 54. 1906. S. 561.

har Opløsningerne paa begge Sider af dette Punkt modsatte elektriske Egenskaber.

Grunden til, at Punktets p_H findes ved 5,2 og ikke ved 6 kunde maaske tænkes at være den, at Tilsætningen af Indicatordraaben frembragte en konstant Forandring i Brintionconcentrationen af Opløsningerne, idet disse jo er stærkt fortyndede og ganske mangler Stødpuder.

n-Leucin. $CH_3 \cdot CH_2 \cdot CH_2 \cdot CH_2 \cdot CH\,(NH_2) \cdot COOH$ undersøgtes i Concentrationen $^n/_{100}$ og med Phosphorsyre og Natron. Dets Ballogram var omtrent identisk med Leucinets.

Phenylalanin. $C_6H_5 \cdot CH_2 \cdot CH\,(NH_2) \cdot COOH$.

Dette Stofs Ballogram fremviser ligesom Leucinets et Punkt, der sandsynligvis maa tydes som det isoelektriske. Det undersøgtes i Concentrationen 0,01 n med Saltsyre og Natron, med og uden Alkohol.

Tabel 14. 0,01 *n* Phenylalanin.

	uden Alkohol		m. 0.2 *n* Alkohol	
$10^4 \cdot C_{HCl}$	Balloel.	p_H	Balloel.	p_H
0	19,7	6,2	52	6,2
1	13,7	4,9	38	4.9
2	5.9	4,55	20,6	4,55
3	÷ 1	4,3	5,1	4,55
4	÷ 3,4	4,2	÷ 5	. .
5	÷ 5,2	.	÷ 11,2	.
10	÷ 4.1	. .	÷ 21.3	3,7
15	0	3,37	÷ 12	3,37
20	2,2	3,22	÷ 6,7	3,22

Det første Skæringspunkt med Nullinien ligger ved $p_H = 4{,}33$ for Forsøgsrækken uden Alkohol, og med Alkohol

ligger det ogsaa meget nær herved. Det isoelektriske Punkt for Phenylalanin ligger som bekendt ved 4,5[1].

Tilsætning af Natron virker omtrent paa samme Maade som ved Leucinet, idet der ogsaa paa den alkaliske Side bliver et Skæringspunkt med Nullinien (ved 0,001 *n NaOH*). C_H her blev ikke bestemt.

Endvidere undersøgtes følgende Ampholyter, der gav mindre interessante Resultater:

Asparaginsyre: $COOH \cdot CH(NH_2)CH_2 \cdot COOH$, undersøgtes i $^n/_{100}$ Opløsning. Den gav hverken alene eller med Saltsyre eller Natron noget Udslag i Ballometret.

Glutaminsyre: $COOH \cdot CH(NH_2)CH_2)_2 \cdot COOH$ undersøgtes i $^n/_{100}$ Opløsning. Den gav intet Udslag, hverken alene eller ved Tilsætning af Phosphorsyre, derimod gav den med Natron et ringe Udslag, der naaede sit Maximum (2,7) ved Neutralisationspunktet af Syren og derefter atter aftog.

α-Amino-Adipinsyre: $COOH \cdot CH(NH_2)(CH_2)_3 \cdot COOH$ undersøgtes i 0,025 *n* Opløsning. Den gav baade alene og med Saltsyre og Natron et ringe Udslag paa omtrent 2.

Lysinchlorid: $NH_2(CH_2)_4 \cdot CH(NH_2) \cdot COOH + 2\,HCl$ undersøgtes i 0,025 *n* Opløsning, men gav intet Udslag, hverken alene, eller ved Tilsætning af Base.

Asparagin: $NH_2 \cdot CO \cdot CH(NH_2) \cdot CH_2 \cdot COOH$ undersøgtes i 0,01 *n* Opløsning. Den gav et Udslag paa 8, der langsomt aftog ved Tilsætning af baade Syre og Base.

Serin: $CH_2(OH) \cdot CH(NH_2) \cdot COOH$ undersøgtes i 0,01 *n* Opløsning. Den gav et Udslag paa 10,4, der langsomt aftog ved Tilsætning af Phosphorsyre eller Natron.

α-Amino-δ-oxy-Valerianesyre. $CH_2(OH)(CH_2)_2 CH(NH_2) \cdot COOH$ undersøgtes i 0,01 *n* Opløsning med Salt-

[1] L. Michaelis. Biochem. Zeitschr. 47. 250. 1912.

syre og Natron og med og uden Alkohol. Alene havde det Balloelektriciteten 19. Ved Tilsætning af Saltsyre sank Udslaget langsomt, indtil det blev 0 omtrent ved $p_H = 1{,}7$ (0,02 *n HCl*); ved Tilsætning af Natron sank det langsomt paa ganske lignende Maade og blev 0 ved omtrent $p_H = 12$ (0,02 *n NaOH*). En Alkoholtilsætning paa 0,2 *n* gav en ganske tilsvarende, men højere liggende Kurve, hvis højeste Punkt svarede til den neutrale Aminosyreopløsning og gav Udslaget 39.

Acetoxim: $(CH_3)_2 \cdot C : N \cdot OH$ undersøgtes i 0,01 *n* Opløsning. Dette gav Udslaget 12,2, som sank langsomt ved Tilsætning af Saltsyre, indtil et Minimum med Balloelektriciteten 4,25 naaedes ved $p_H = 4{,}05$, hvorefter Udslaget atter blev større. Udslaget forøgedes paa hele denne Strækning baade ved Tilsætning af 0,2 *n* Alkohol og af 0,1 *n* Natriumchlorid til henved det dobbelte. Tilsætning af Natron forøgede Udslaget til omtrent 17 (ved $p_H = 9{,}3$). Paa den alkaliske Side havde Alkohol forøgende, Natriumchlorid formindskende Virkning paa Udslaget; ved et givet Punkt, nemlig $p_H = 6{,}6$ var Salttilsætning uden Virkning.

Ogsaa en stærkere Concentration af Acetoxim undersøgtes, nemlig 0,05 *n*. I denne Concentration virkede 0,1 *n NaCl* stærkt forøgende paa Udslaget baade paa sur og alkalisk Side.

Amino-Benzoesyre. $C_6H_4\,(NH_2) \cdot COOH$. Baade *o*, *m* og *p* prøvedes, men gav alle kun minimal balloelektrisk Virkning, baade alene og med Syre eller Base.

Theobromin, $C_5H_2\,(CH_3)_2 \cdot N_4O_2$ undersøgtes i Concentrationerne 0,0017, 0,005 og 0,01 *n* med Saltsyre og med og uden Alkohol. Det balloelektriske Udslag var for alle disse Concentrationer i ren Opløsning omtrent 16, og det sank ved Tilsætning af Saltsyre stærkt, indtil Kurven skærer Nullinien omtrent ved $p_H = 4{,}05$ for alle Concentrationer.

hvorefter Udslaget blev negativt. Alkoholtilsætning virkede ens baade før og efter dette Neutralpunkt, idet Udslaget ved alle Brintionconcentrationer (undersøgt mellem $p_H = 6{,}5$ og 3,7) var stærkt positivt, men alle tre Alkoholkurver viste et udtalt Minimum ved $p_H = 4$.

Ifølge Litteraturen ligger det isoelektriske Punkt for Theobromin ved $p_H = 5{,}2$[1].

Kakodylsyre, $(CH_3)_2\ As O \cdot OH$: 0,01 n Kakodylsyre havde $p_H = 4{,}74$ og gav i sig selv Balloelektriciteten 10. Ved Tilsætning af Saltsyre aftog Udslaget, indtil Kurven skar Nullinien ved $p_H = 3{,}85$, hvorefter Udslagene blev negative. 0,02 n Kakodylsyre, blandet med 0,02 n Saltsyre, var balloelektrisk neutral ($p_H = 2{,}35$) ligesom ogsaa en endnu surere Opløsning ($p_H = 1{,}9$).

Alkoholtilsætning bevirkede her ligesom ved Theobromin og Theophyllin positivt Udslag baade før og efter Skæringspunktet ($p_H = 3{,}85$) af den alkoholfri Kurve, men med et Minimum svarende til dette.

Kakodylsyrens isoelektriske Punkt ligger ved $p_H = 3{,}9$[2], altsaa meget nær ved det her omtalte Punkt.

Stærkere Kakodylsyreopløsninger (0,05 n og 0,025 n) var i sig selv svagt negativt balloelektriske og forandredes kun lidt ved Saltsyretilsætning.

Slutning.

Det vigtigste Resultat af de foreliggende Undersøgelser er, at visse Ampholyters isoelektriske Punkt synes at være neutralt i balloelektrisk Henseende, mens dette Punkt for andre Ampholyters Vedkommende ikke kan findes paa Ballogrammet.

Da Serumalbumin og Ægalbumin i udpræget Grad viser

[1] L. Michaelis u. H. Davidsohn. Biochem. Zeitschr. 30. 143. 1910. Interessant er det, at Forf. finder Krystallisationsoptimum ved $p_H = 4$.

[2] Zeitschr. f. physik. Chemie 54. 1906. S. 561.

dette Punkt, og da det fra W. Paulis[1] Undersøgelser er bekendt, at disse Stoffers Joner er stærkt hydratiserede, er det rimeligt at antage, at det er Hydratiseringen af Jonerne, der gør dem balloelektrisk virksomme. Hvis denne Hydratisering indenfor samme homologe Række stiger med Molekylvægten, — analogt med Alkoholers og Fedtsyrers Hydratdannelse[2], — bliver det forstaaeligt, at Glycin- og Alanin Jonerne er balloelektrisk uvirksomme, medens Amino-Smørsyre, Amino-Valerianesyre og Leucin frembyder Ballogrammer med karakteristiske Punkter, der med megen Sandsynlighed kan tydes som isoelektriske.

Efterskrift af Johanne Christiansen.

Forsøgene i ovenstaaende Arbejde er udført allerede i 1914 af min Fader og mig i Forening. Arbejdet blev dengang foreløbig henlagt, da de Forsøgsresultater, vi opnaaede, var vanskelige at tyde, og da vi haabede senere at blive istand til at kunne fortolke dem, naar først visse andre Sider af Balloelektriciteten var opklarede. Navnlig haabede min Fader, at det af ham i 1916—17 konstruerede Kondensatorballometer vilde være en stor Hjælp ogsaa ved Arbejdet med amphotere Stoffer. Ved min Faders Død i 1917 forhindredes Planerne om at vende tilbage til dette Arbejde i Fællesskab, og jeg har derfor anset det for rigtigt at udgive det foreliggende Forsøgsmateriale med en kortfattet Text, da det, selv om det er ufuldstændigt, dog er værdifuldt som Grundlag for fortsatte balloelektriske Forsøg med disse Stoffer.

[1] l. c.

[2] Johanne Christiansen. Zeitschr. f. physikal. Chemie 92. 257. 1917.

[3] C. Christiansen. Experimentalundersøgelser over Gnidningselektricitetens Oprindelse VIII. Vid. Selsk. math.-fysiske Meddel. II, 7. 1919.

Forelagt ved Prof. S. P. L. Sørensen paa Mødet d. 14. Febr. 1919.
Færdig fra Trykkeriet d. 11. December 1919.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. II, 7.

EXPERIMENTALUNDERSØGELSER OVER GNIDNINGSELEKTRICITETENS OPRINDELSE. VIII

AF

C. CHRISTIANSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HOST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1919

Efter Professor CHRISTIANSENS efterladte Manuskript, færdig redigeret af JOHANNE CHRISTIANSEN.

§ 1. Indledning. Naar to isolerede Legemer gnides mod hinanden og derefter adskilles, vil man vistnok altid finde, at de have modsatte elektriske Ladninger. Som oftest ere disse Ladninger fastsiddende, men der er dog ogsaa en Mulighed for, at en Del af den frembragte Elektricitet er gaaet ud i Luften; i saa Fald behove de nævnte Legemer ikke at have ligestore men modsatte Ladninger. Under almindelige Forhold vil det være vanskeligt at opsamle og maale den Elektricitetsmængde, som er gaaet over i Luften; derimod er det meget let at gjøre det ved at gaa ud fra det af mig tidligere beskrevne Ballometer.

Inden jeg gaar over til at beskrive mine Forsøg herover skal jeg i Korthed gjøre Rede for de vigtigere Arbejder i samme Retning, som findes i Litteraturen. J. J. THOMSON[1] ledede en Luftstrøm gjennem Vand; at Luften derved blev ioniseret viste sig ved, at den afledede et ladet Elektroskop hurtigere end almindelig Luft. Virkningen var den samme enten Elektroskopet var ladet med positiv eller negativ Elektricitet. Lignende Virkning gav vandige Opløsninger. Derimod vare Æther, Alkohol og Terpentin uvirksomme.

[1] J. J. THOMSON. Phil. Mag. (4) T. 6. S. 352. 1902.

K. Kähler[1] undersøgte de Joner, som faldende Vanddraaber frembringe i Luften. De vare alle negativt ladede og havde en Vandringshastighed af 4,17 Centimeter i Sekundet, naar den elektriske Kraft var 1 Volt/cm. Forsøg med en Kogsaltopløsning, som selv blev negativ, gav derimod det Resultat, at der i Luften fandtes Joner af begge Slags, de positive havde en meget lille Vandringshastighed, de negative forholdt sig omtrent som de Joner, Vandet giver.

E. Aselmann[2] har senere behandlet den samme Opgave; han har i Almindelighed fundet de samme Resultater som Kähler, men er gaaet dybere ind paa Sagen. Han finder da ogsaa, at destilleret Vand kun afgiver negative Joner til Luften; for de fleste af dem ligger Vandringshastigheden mellem

$$v = 4{,}0 \text{ og } v = 1{,}6 \cdot 10^{-2}.$$

Men der findes ogsaa nogle, for hvilke den ligger imellem

$$v = 1{,}6 \cdot 10^{-2} \text{ og } 2{,}7 \cdot 10^{-4}.$$

En 0,2 % Natriumchloridopløsning giver begge Slags Joner. Af de positive har de fleste en Vandringshastighed mellem

$$n = 8{,}8 \cdot 10^{-2} \text{ og } n = 6{,}06 \cdot 10^{-4}$$

Resten ligger mellem

$$n = 6{,}06 \cdot 10^{-4} \text{ og } n = 3{,}4 \cdot 10^{-4}$$

De negatives Vandringshastighed ligger mellem

$$v = 4{,}0 \text{ og } v = 1{,}94 \cdot 10^{-3}.$$

Man kunde tænke sig, at de langsomme og derfor tunge Joner vare fremkomne ved Fortætning af Vanddamp; dette er efter Aselmann ikke Tilfældet; heller ikke bestaa de af

[1] Annalen der Physik. Bd. 12. S. 1119. 1903.
[2] Annalen der Physik. Bd. 19. S. 960. 1906.

Draaber af Vædsken; de maa altsaa antages at bestaa af Samlinger af Luftmolekyler.

Paa en meget slaaende Maade har A. S. Eve[1] viist, at der opstaar Joner af begge Slags i stor Mængde, naar Vædsker omdannes til meget smaa Draaber. Han sendte en kraftig Luftstrøm forbi Mundingen af et snevert Rør, som stod ned i Vædsken, der skulde undersøges. I den Blanding af Luft og Vædskestøv, som derved dannes, fandt han store Mængder af begge Slags Elektricitet. Jeg gjengiver her en stor Del af hans Resultater.

	K	A	B
Kvægsølv	0	0	0
Toluen	0,02	0,02	0
Rhigoline	0,03	0,07	+ 0,04
Benzin	0,08	0,11	+ 0,03
Phinetol	0,13	0,13	0
Terpentin	0,11	0,04	− 0,07
Svovlkulstof	0,07	0,07	0
Saltvand	0,015	0,01	0,5
Natrium Carbonat opl.	..	0,07	..
Saltsyre	0,04	0,04	0
Ammoniakvand	0,45	0,67	+ 0,22
Vandværksvand	0,50	0,62	+ 0,12
Destilleret Vand	1,0	1,6	+ 0,6
Amylacetat	1,9	1,9	0
Acetone	2,4	2,4	0
Aldehyd	2,5	2,5	0
Chloroform	2,7	2,7	0
Eddikesyre	3,7	3,7	0
Methyljodid	3,7	3,7	0
Methylalkohol	4,0	4,0	0
Amylalkohol	4,5	4,5	0
Æthylalkohol	4,6	4,6	0
Æther	3,7	5,0	+ 1,3

Under K er anført Mængden af positiv, under A Mængden af negativ Elektricitet og under B den Elektricitets-

[1] A. S. Eve Phil. mag. Vol. XIV, p. 383. 1907.

mængde, som Vædsken har optaget i sig. B er altsaa Balloelektriciteten. Heraf fremgaar det vigtige Resultat, at de allerfleste rene vandfrie Vædsker frembringe lige meget Elektricitet af begge Slags, den eneste Undtagelse fra denne Regel er Æther. Vand og vandige Opløsninger blive i Reglen positivt elektriske, hvilket ogsaa stemmer med hvad Lenard har fundet.

De saaledes dannede Joner forbinder sig efterhaanden med hverandre. Hurtigst sker dette med Vandets Joner. Den Hastighed, hvormed Gjenforeningen finder Sted, aftager, naar man gaar fremad i Rækken.

Vand
Amylalkohol
Æthylalkohol
Eddikesyre
Æther
Benzylalkohol

For Vandets Vedkommende forandredes Elektricitetsudviklingen ikke kjendeligt, selv om det opvarmedes til Kogepunktet.

§ 2. Kondensatorballometret. Naar Draaberne i Ballometret støde mod den vædede Platinplade, vil en vis Mængde Elektricitet B blive i Ballometret, det er den som man maaler ved Hjælp af det med Ballometret forbundne Elektrometer. B er da det, som jeg har kaldt Balloelektriciteten eller kortere Balliteten. En tilsvarende Elektricitetsmængde $\div B$ føres da ud af Ballometret med Luftstrømmen, som forlader det. I hvilken Form $\div B$ er tilstede er det vanskeligt at vide, den kan tænkes at bestaa af frie Joner, ladede Molekyler eller ladede Vædskedraaber. Det vil være rigtigst at betragte de positive og de negative Ladninger hver for

sig. De positive Ladninger kalder jeg Kationiteten, de negative Anioniteten og betegner dem henholdsvis med $+K$ og $-A$. Man maa da have

$$-B+K-A=0.$$

Hensigten med Kondensatorballometret er da at maale K og A. Dette kan ske ved at lede Luftstrømmen ind i Rummet mellem en stor Luftkondensators to Plader. Lader man den ene af Kondensatorpladerne C_1 til en tilstrækkelig høj positiv Spænding, vil hele Anioniteten A trækkes hen til denne, medens Kationiteten K drives over paa den anden Kondensatorplade C_2; den kan da maales paa sædvanlig Maade ved Elektrometret. Lades C_1 derimod med negativ Elektricitet, kan Anioniteten A maales.

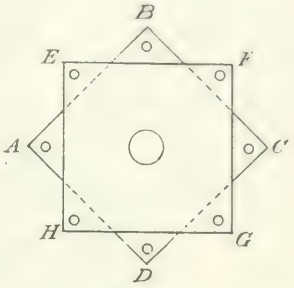


Fig. 1.

Kondensatoren lavedes paa følgende Maade. Den bestod af 19 Zinkplader, 16 cm i Kvadrat, med et rundt Hul, 2 cm i Kvadrat, i Midten. I Hjørnerne af den underste Plade *ABCD* (Fig. 1) befæstedes 4 Messingrør, noget over 20 cm lange. Den næste Plade *EFGH* anbragtes i den i Figuren viste Stilling, den var forsynet med lodrette Messingrør ligesom den første Plade. Man lader nu 4 Messingrør, som ere 2 cm lange, og som netop kunne gaa uden om de lange Messingrør, glide ned om disse. Derpaa lader man den tredie Zinkplade, som har fire Huller i Hjørnerne, glide ned ad de paa *ABCD* anbragte Stolper, de vil da lægge sig paa de omtalte korte Messingrør. Derpaa gjøres det samme for *EFGH*'s Vedkommende og paa denne Maade fortsættes. Tilsidst lukkes Hullet i Midten af den øverste Plade. Den underste Plade *ABCD* stilles paa et isolerende

Underlag, den næste *EFGH* paa 4 Ravplader, saaledes at der bliver en Afstand af 1 cm mellem de to Plader.

Kondensatorballometret selv er fremstillet skematisk i Gjennemsnit i Fig. 2. *AA* er en Kapillaraspirator som under Paavirkning af en Luftstrøm, der ledes ind gjennem Røret *D*, suger Vædsken, der skal undersøges, op af Skaalen *F*, der staar paa et isolerende Underlag, og sprøjter den imod en Platinplade *G*. *AA* bæres af en Glasbeholder *HH*. Vædskedraaberne som falde ned fra *G* samles i Bunden af *HH* og gaa derfra gjennem det snevre Rør h_1 ned i en Skaal K_1. Kapillaraspiratoren 'virker dog desværre ikke altid som den skulde. Ikke sjældent seer man, at noget af Vædsken ikke deltager i Sprøjtningen men derimod drives ned ad Undersiden af Kapillaraspiratoren og gjennem det snevre Rør h_2 gaar ned i Karret K_2. Denne »Savlen« medfører den store Ulempe at man ikke kan sammenligne to Forsøg med hinanden, med mindre man har overbeviist sig om, at den Mængde Vædske som er forvandlet til Draaber, i begge Tilfælde er den samme. Jeg har dog ment at kunne hjælpe derpaa ved følgende Beregning. Reduceret Balloelektricitet B_0 kalder jeg Forholdet

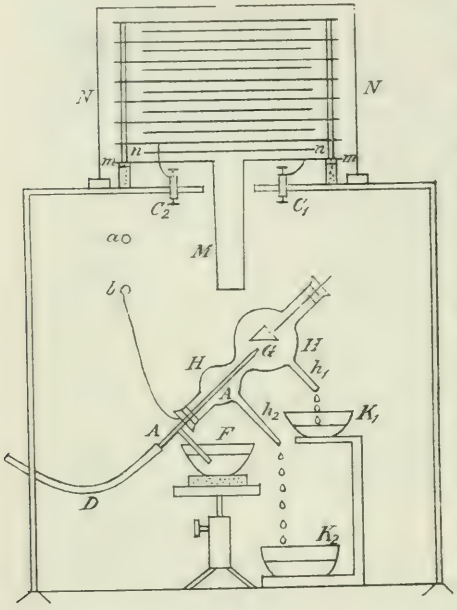


Fig. 2.

$$B_0 = \frac{B}{V}$$

mellem Udslaget B paa Elektrometret og Rumfanget V af den Vædske, som i Tidsenheden opsamles i Karret K_1, forudsat at Lufttrykket bestandig er det samme.

Den ved Stødet ioniserede Blanding af Luft og Vædskedraaber ledes gjennem Metalrøret M, som er loddet til den underste Kondensatorplade mm, og udbreder sig herfra over hele Kondensatoren. Fra Pladen mm selv fører en isoleret Ledning hen til C_1, som efter Omstændighederne forbindes med den ene eller den anden af et Akkumulatorbatteris Poler.

Ballometret selv staar nede i en Kasse, dannet af perforerede Zinkplader, som er forsynet med en Skydedør af samme Materiale. Oven paa denne Kasse ligger en Træplade, som bærer Kondensatoren. Denne er ved en omvendt Zinkbeholder NN beskyttet mod ydre elektriske Kræfter og mod Luftstrømninger. NN er i ledende Forbindelse med Pladen mm og derigjennem med Akkumulatorbatteriet.

Med Hensyn til Brugen af Kondensatorballometret bemærkes følgende:

1. Maaling af Balliteten B og B_0.

Elektrometret ene Pol E, der er i Forbindelse med Ballometret, er ved en Shunt med meget stor Modstand forbundet med Jorden. Denne Shunt dannes af et eller flere Haarrør fyldte med Nitrobenzol. Den anden Pol er forbundet med Jorden. Naalen er ladet med Byens Spænding (220 Volt). Lad Ledningen fra E til Ballometret ende ved a. En Kobbertraad fører fra Ballometrets Indre hen til b. Forbindes nu a og b med hinanden, saa vil Elektrometret gjøre et Udslag, der, naar Elektrometrets Udslag er proportionale med Potentialforskellen mellem dets Poler, er et Maal for Balloelektriciteten. Vil man have den reducerede Ballitet B_0, maales den Vædskemængde, som i given Tid falder ned i Skaalen K_1.

2. Maaling af Joniteterne K og A, K_0 og A_0.

Kationitet K kalder jeg den Mængde positiv Elektricitet, som Luftstrømmen fører med sig, naar den forlader Ballometret. Den maales ved at forbinde C_1 og dermed alle Kondensatorplader af ulige Orden med Akkumulatorbatteriets positive Pol, medens C_2 og dermed nn og alle de andre Kondensatorplader af lige Orden forbindes med Elektrometret; samtidig forbindes Ballometret ved en Ledning fra b med Jorden. Paa lignende Maade findes A, K_0 og A_0; de to sidste, de reducerede Joniteter, svare til den reducerede Ballitet B_0.

§ 3. Eksempler paa Brugen af Kondensatorballometret. Vi skulle først see, hvorledes den tidligere omtalte Kontrolblanding, bestaaende af lige Rumfang $2m$ Aethylalkohol og $\frac{1}{10}n$ Natriumchlorid forholdt sig i Ballometret ved forskellige Tryk.

Tabel 1. 20 Akkumulatorer.

Ballolyt	Tryk	C	B	K	A	S
Kontrol	6,0 cm Hg	÷ 2,5	12,5	14	÷ 24	2,5
Kontrol	7,4 - -	÷ 5	25	22	÷ 48	+ 1
Kontrol	9,2 - -	÷ 10	50	44	÷ 94	0

Her er B det Udslag, som fremkommer, naar Ballometret direkte er forbundet med Elektrometret, altsaa Balliteten selv; K og A Udslagene, naar den ene Halvdeel af Kondensatoren er forbundet resp. med Akkumulatorbatteriets positive eller negative Pol, medens den anden Halvdeel er forbundet med Elektrometret, K er altsaa det, jeg i det foregaaende har kaldt Kationiteten, A derimod Anioniteten. S er Summen af B, K og A; hvis Apparatet var fuldkomment

skulde S være lig 0. Naar S ikke er Nul, kan der være flere Aarsager dertil. Fremfor alt gjælder det om, at al den Elektricitet, som kommer ind i Kondensatoren, ogsaa optages af den; om dette skeer, afhænger dels af Akkumulatorbatteriets Spænding, dels af Jonernes Vandringshastighed. C er det Udslag, som fremkommer, naar den ene Halvdeel af Kondensatoren er afledet til Jorden, den anden er forbundet med Elektrometret. C er, som man kunde vente, proportional med B.

Af Tabellen i sin Helhed fremgaar det vigtige Resultat at Forholdet mellem Ballitet og Jonitet er uafhængig af Trykket.

Som Eksempel paa, hvorledes Akkumulatorspænding og Tryk influerer paa Apparatets Angivelse, skulle nogle andre Forsøg anføres.

Tabel II.

Ballolyt	Tryk	Akk.	B	K	A	S
Kontrol	7,5	1	25	3,7	÷ 17,7	11,0
Kontrol	7,5	2	25	7,2	÷ 26,5	5,7
Kontrol	7,5	3	25	9	÷ 32,3	1,7
Kontrol	7,5	6	25	9	÷ 33,3	0,7
Kontrol	9,5	6	48	15	÷ 60	3

I alle disse Forsøg har Akkumulatorspændingen været for lille; for at finde de virkelige Værdier af K og A, som vi kalder K_1 og A_1, kunde man sætte (af Forsøgene i Tab. II).

$$K_1 + A_1 + B = 0$$

$$K_1 = 9\ (1 + x), \quad A_1 = -33{,}3\ (1 + x)$$

som giver

$$K_1 = 9{,}3, \qquad A_1 = -34{,}3.$$

Herefter vilde de tabte Elektricitetsmængder være

	$K_1 - K =$	$A_1 - A$
1 Akk.	5,6	— 16,6
2 Akk.	2,1	— 7,8

At Elektricitetstabet blev større i det sidste end i det næstsidste Forsøg var at vente, da Trykket og altsaa ogsaa Luftstrømmens Hastighed der er større.

Tabel III.

Ballolyt	Tryk	Temp.	*B*	*C*	*K*	*A*	*S*
Vandværksvand....	9,2	15—16°	÷ 1	÷ 0,5	23	÷ 23	÷ 1
Destilleret Vand...	9,2	15—16°	8,5	÷ 1	21,5	÷ 29	+ 1
Ledningsevnevand..	9,2	15—16°	9	÷ 1	22,5	÷ 31	+ 1,5

At Vandværksvand bliver negativt ved Stød mod sig selv, forstaas let, da det væsentlig er at ansee for en Saltopløsning, og disse blive, som det er viist tidligere (III. S. 233), altid negativt elektriske. I den Anledning har jeg anstillet følgende Betragtning. Naar Glas gnides med en Linnedklud, bliver det i Reglen positiv elektrisk. Men det skeer dog ogsaa, at det bliver negativt elektrisk. Det faldt mig ind, at dette kunde afhænge af, om Glasset var renset med Vandværksvand eller med destilleret Vand. Jeg har gjort nogle Forsøg derover med Rør af Jenaglas og finder at det virkelig forholder sig saaledes.

Tabel IV.

Ballolyt	Tryk	Temp.	*B*	*C*	*K*	*A*	*S*
Alkohol I ..	6,0	c. 15	0	÷ 1	25	÷ 27	÷ 2
Alkohol II ..	6.0	c. 15	÷ 1	..	50	÷ 52	÷ 3

Alkohol I var Alkohol som den faas i Handelen under Betegnelsen absolut Alkohol; Alkohol II var 99% Alkohol fra Petri. At Alkohol forholder sig paa en meget lunefuld Maade i ballolytisk Henseende, er for længe siden bemærket af P. Lenard (III S. 215)

Ved Kondensatorballometrets Hjælp see vi altsaa, at der slet ingen Elektricitet udvikles, og dog forholder Sagen sig ganske modsat. Men selv i de Tilfælde, i hvilke der finder en stærk Elektricitetsudvikling Sted som ved Blandingen af Salte med Alkohol er dog den Elektricitetsmængde, der gaar over i Luften, langt større end den, der bliver tilbage i Ballometret. Alt dette har dog kun Betydning som en Bekræftelse paa det forlængst af Eve fundne Resultat.

Maaling af Vandringshastigheden. Ved Vandringshastighederne u og v forstaar man den Hastighed som resp. den positive og den negative Jon modtager under Paavirkning af Enhed af elektrisk Kraft, hvorved vi her ville forstaa Volt pr. Centimeter. Til at maale dem ligger det her nærmest at benytte en Metode, som er angivet af Rutherford[1]. Man tænke sig to conaxiale Cylindre, den yderste ubegrændset, dens Radius er a, den inderste har Længden l og Radius b. Deres Potentialer være resp. V og Nul. Den elektriske Kraft i Rummet mellem Cylindrene er da

$$X = -\frac{V}{x \ln \frac{a}{b}}$$

idet x er Afstanden fra Cylindrenes fælles Axe. Kaldes en positiv Jons Hastighed U, har man da

$$U = -\frac{Vu}{x \ln \frac{a}{b}}.$$

Til Bestemmelse af Banen har man nu

$$dx = -\frac{Vu}{x \ln \frac{a}{b}} \cdot dt, \quad dy = c dt$$

[1] Phil. mag. Vol. 47. S. 146. 1899.

idet c er Luftstrømmens Hastighed. Har Jonen fra først af Afstanden r fra Axen, faas heraf

$$r^2 - x^2 = \frac{2\,Vul}{ln\frac{a}{b}}, \quad y = ct.$$

Lad Tætheden af den positive Elektricitet i Luftstrømmen være ρ, saa vil den Mængde positiv Elektricitet E, som rammer den indre Cylinder i Tidsenheden være

$$E = \pi(r^2 - x^2)\rho c = \frac{2\pi V\rho ul}{ln\frac{a}{b}}.$$

Tillige haves den hele Mængde positiv Elektricitet, som i Tidsenheden strømmer igjennem Røret lig

$$K = G\rho$$

altsaa er

$$\frac{E}{VK} = \frac{2\pi ul}{G\,ln\frac{a}{b}}.$$

naar G er det Rumfang Luft, som gaar ud af Ballometret i Tidsenheden,

eller

$$u = \frac{E}{VK} \cdot \frac{G \cdot ln\frac{a}{b}}{2\pi l}.$$

For i Praxis at kunne udføre disse Maalinger, anbragtes inde i Røret M (Fig 3) (Lgd. 18,6 cm indv. Diam: 19 mm.) der staar i Forbindelse med den ladede Halvdel af Plade-Kondensatoren (se Fig. 2), et i begge Ender lukket Messingrør, 10,3 cm langt og med udvendig Diameter: 6 mm. Dette Rør kaldes Mikrokondensatoren. Mikrokondensatoren er fuldstændig isoleret fra M paa følgende Maade: Udenom M er anbragt en Ebonit-Ring, E (se Fig. 3), hvorom igen ligger en Messingring, der paa den ene Side bærer en Messingplade. p. paa den anden Side en Kontakt, k, ved

Hjælp af hvilken Ringen sættes i Forbindelse med Jorden. Hensigten hermed er at hindre, at der gaar Elektricitet over fra den stærkt ladede Halvdel af Plade-Kondensatoren der jo er i ledende Forbindelse med M, til Mikrokondensatoren. Pladen, p, er ved Hjælp af et 6 mm tykt isolerende Lag af Glas, beklædt med Lak, adskilt fra en anden Plade p_1, der ved Hjælp af Stilken a, der gaar gennem Midten af et Hul, o, i M, bærer Mikrokondensatoren.

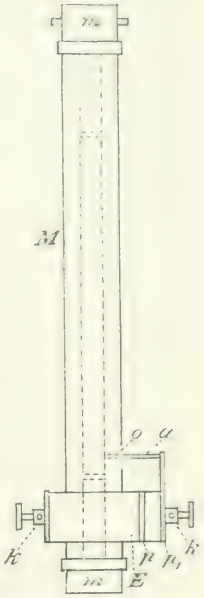


Fig. 3.

p, og dermed Mikrokondensatoren kan ved Hjælp af Kontakten k_1, sættes i Forbindelse med Elektrometret.

For hver Ende af M er anbragt et lille Mellemstykke af Messing, m, der kan skydes ind i M, og som ved Hjælp af tre smaa Egere bærer et Messingrør af samme Kaliber som Mikrokondensatoren. Disse to Rør ender faa mm henholdsvis over og under Mikrokondensatoren og tjener til at fordele Luftstrømmen. Den øverste Messingring kan ved en Bajonetlaas sættes i Forbindelse med Plade-Kondensatoren. Den nederste Ring kan ved Hjælp af en Prop og en vid Glashane sættes i Forbindelse med Ballometret.

Denne Glashane tjener til fuldstændig at aflukke Ballometret foroven, hvilket gjores, naar man ønsker at maale den Luftmængde, der passerer Ballometret i Minuttet (Størrelsen G i Beregningen). I saa Tilfælde anbringer man et Rør i den Prop, der bærer Ballometrets Platinplade, og sætter det ved Hjælp af en Kautchoukslange i Forbindelse med et Gasur.

Dersom alle de frembragte Joner af et og samme Fortegn havde samme Vandringshastighed, saaledes som forudsat i den ovenstaaende Beregning, vilde denne simpelt kunne bestemmes ved det beskrevne Apparat. Man vilde da blot behøve, dels en Maaling ved forholdsvis høj Spænding af den samtlige Mængde Joner af det paagældende Fortegn, dels en Maaling af den Brøkdel af Ladningen, der opfanges, naar Røret udenom Mikrokondensatoren er ladet til en ringere Spænding (V). Derved vilde man nemlig bestemme Størrelserne K og E, og ved Hjælp af Formlen nederst S. 14 kan derfor u bestemmes, naar man kender Mikrokondensatorens Dimensioner, Luftstømmens Hastighed (G) og Spændingen (V).

Man maa imidlertid ved Forsøgene være forberedt paa at finde, at ikke alle Jonerne har samme Vandringshastighed og for nærmere at undersøge, hvordan Jonerne kan opdeles i Grupper med forskellig Vandringshastighed, vil Maalinger af den i Mikrokondensatoren opfangede Elektricitetsmængde for forskellige Spændinger være nødvendig.

Efterskrift af Johanne Christiansen.

Professor Christiansen blev ved sin Død 1917 afbrudt midt i Udarbejdelsen af denne Undersøgelse. Medens det, som det fremgaar af denne Afhandling, til hvilken Manuskript forelaa fra Professor Christiansens Haand, var lykkedes ham at faa en Del Resultater vedrørende de absolute, positive Balliteter, saa fik han med Mikrokondensatoren kun gjort ganske enkelte Forsøg.

Det havde været hans Hensigt i større Forsøgsrækker navnlig over Blandinger af Opløsninger af Elektrolyter og Ikke-Elektrolyter, at bestemme saavel de to Slags Balliteter

som de forskellige Joners Hastigheder, men denne Plan, som han ventede sig meget af, og som sikkert i hans Hænder vilde have ført til interessante Resultater, maa nu ligge hen, indtil — hvad der var hans største Ønske — yngre Fysikere eller Kemikere faar Lyst til at tage Sagen op.

Nærværende Undersøgelse er ligesom No. V, VI og VII foretaget paa Universitetets Institut for almindelig Pathologi, hvor min Fader de sidste fire Aar af sit Liv ved Professor C. J. Salomonsens Velvillie fandt udmærkede Arbejdsbetingelser.

Forelagt paa Mødet d. 2. Nov. 1917.
Færdig fra Trykkeriet d. 11. December 1919.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. II, 8.

OVERFLADESPÆNDINGENS INDFLYDELSE VED UDSTRØMNING AF EN VÆDSKE I STRAALEFORM

AF

JUL. HARTMANN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HOST & SON, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1919

INDLEDNING

En Indflydelse af Overfladespændingen paa Udstrømningen af en Vædske i Straaleform har længe været kendt. Men den Maade, hvorpaa Spændingen gør sig gældende, har man ofte i høj Grad misforstaaet. Naar der af et og samme Straalehul ved samme Trykhøjde pr. Sek. udstrømmer et større Volumen af en Vædske med en mindre end af en med større Overfladespænding, har man saaledes villet forklare dette ved, at Overfladespændingen i sidste Tilfælde snører Straalen mere sammen og altsaa forøger Kontraktionen. Man finder f. Eks. denne Forklaring i J. Jamins Lærebog »Cours de Physique« Bind I. Maaske er den herfra eller fra J. Jamins Kilde gaaet over i andre Lærebøger, hvor den ogsaa optræder. Jeg gengiver Jamins Fremstilling:

»Girard[1] avait déjà constaté que l'alcool coule plus rapidement que l'eau à travers des orifices percés en mince paroi: la section contractée de la veine doit donc être plus grande pour l'alcool que pour l'eau, et c'est en particulier ce qui résulte de mesures directes de M. Isarn[2]. Il suffit, d'après ce dernier savant, de vaporiser de l'alcool ou de

[1] Girard, Mémoires de l'Académie, t. I. 1816.

[2] Isarn, Journal de Physique, Ire série, t. IV, p. 167; 1875.

l'éther à peu de distance d'un orifice par lequel de l'eau s'écoule, pour que, le liquide à faible tension superficielle se disolvant dans l'eau, la dépense de l'orifice augmente dans une proportion appréciable.«

Den, der først har givet den i Følge de nedenfor anførte Forsøg rigtige Forklaring af Overfladespændingens Indflydelse, er vistnok C. Christiansen[1]. Hvis Overfladespændingen er C, maa der i en cylindrisk Straale med Radius r efter Christiansen virke et Tryk C/r, som modsætter sig Udstrømningen, og som bliver at trække fra det Tryk, under hvilket Vædsken strømmer ud, naar man paa Grundlag af Torricellis Formel vil beregne den udstrømmende Mængde. Christiansen prøvede paa at gennemføre de Forsøg, som jeg nu har tilendebragt, men Forsøget mislykkedes og maatte mislykkes paa Grund af Ufuldkommenheden af de Apparater, hvoraf han betjente sig. Jeg genoptog Forsøgene og gjorde det, især fordi de stillede en simpel Metode til Bestemmelse af Overfladespænding i Udsigt, og fordi det paa Forhaand vel ikke kunde betragtes som aldeles givet, at Christiansens Anskuelse om Indflydelsen holdt Stik. Resultatet af mit Arbejde blev en meget fuldkommen Bekræftelse paa Christiansens Antagelse, idet en Bestemmelse paa Grundlag af denne af Overfladespændingen for Kvægsølv og Vand førte til Værdier for disse Konstanter, praktisk set sammenfaldende med de bedste hidtil udførte Bestemmelser. Efter at dette var fastslaaet, har jeg paa Grundlag af de indvundne Erfaringer uddannet en ny Maalemetode til Overfladespændingens Bestemmelse. Beretningen om dette Arbejde er fremkommen andetsteds.[2]

[1] C. Christiansen, Ann. d. Phys. 5, 436, 1901.

[2] Jul. Hartmann, Ingenioren Nr. 53—55, 1919.

Forsøgsapparat.

Dette, der er vist i Fig. 1 *a*—*c* bestod ved mine Forsøg af en cylindrisk udboret Stobejærnskumme med en indvendig Diameter paa 14,845 cm og en Dybde paa ca. 9,5 cm. Paa modsatte Sider havde Beholderen udvendig to »Forstærkninger« *F* med plane Ydersider. Gennem hver af disse og Cylinderens Væg var der ført to Boringer Fig. 1 *c*. Den ene af disse var i hver Forstærkning bestemt til at optage et planparallelt Vindu *V*. I den anden Boring i den ene Forstærkning skruedes det straaledannende Element *P* (Hulpladen). Den tilsvarende Boring i den anden Forstærkning benyttedes ikke og var derfor lukket med en Prop.

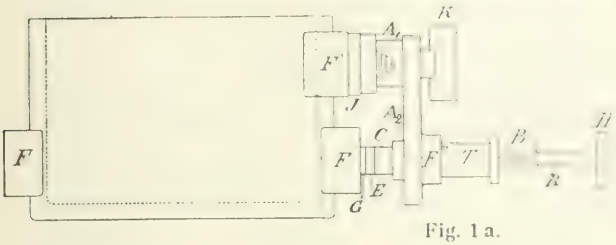


Fig. 1 a.

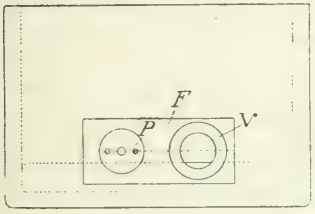


Fig. 1 b

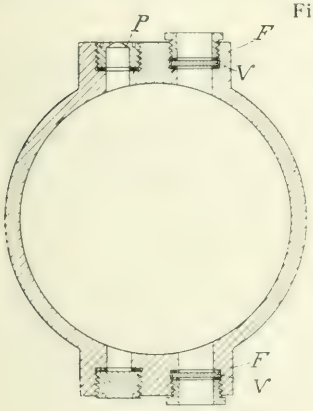


Fig. 1 c.

Ved Forsøgene skulde man saa pludselig som muligt kunne aabne og lukke for Straalehullet. Lukkemekanismen er fremstillet i Fig. 1 a. Den var bygget ud fra den plane Forside af Forstærkningen *F'*. Fra *F'* var den isoleret ved Fiberskiven *I*. Paa denne var den korte, vandrette

Arm A_1 fastgjort. Paa A_1 kunde atter den lodrette Arm A_2 fastspændes ved Fløjskruen K, saaledes at Armen tillige med alt, hvad der sad paa den, let kunde fjærnes, naar Hulpladen skulde skiftes. A_2 havde forneden en Forstærkning F''. I denne kunde den egentlige Laas skydes frem og tilbage og fastspændes i en passende Stilling. Laasen bestod af en Staalskive E paa Enden af den massive Cylinder C. E var langs Randen forsynet med en plan Gummiring, der passede i et neddrejet Spor. Ringen pressedes, naar Laasen var lukket, ind mod den plane Forside af Straalehulspladen. Trykket udøvedes herved af en kraftig Fjeder monteret indeni den hule Cylinder T. Ved at trække Stangen R med Haandgrebet H udad aabnedes der for Vædsken. Ved en Bajonetlaasindretning B kunde Lukkemekanismen efter at være trukket ud stilles i Ro. Endelig kunde den lukkes ved en kort Drejning af H, som det vil forstaas af Figuren.

Naar Laasen isoleredes fra Vædskebeholderen, var det, fordi der ved andre Forsøg end dem, der her er Tale om, maatte være Mulighed for at sende en elektrisk Strøm gennem Straalen. Skiven E skulde herved tjene som den ene Elektrode.

Med det beskrevne Forsøgsapparat arbejdedes der paa følgende Maade. I Kummen hældtes der først 250 cm^3 af den Vædske, med hvilken Udstrømningsforsøget skulde udføres. Vædskens Overflade dannede en Bund i Kummen, og denne Bunds Dybde under Centrum af Straalehullet udmaaltes ved et Kathetometer med en Nøjagtighed af ca. 0,05 mm. Herefter hældtes der yderligere f. Eks. 1250 cm^3 Vædske i Kummen. Idet dennes indvendige Diameter var udmaalt — ligeledes ved Kathetometer — kunde Vædskeoverfladens Højde over Straalehullets Centrum beregnes.

Udstrømningsforsøget bestod nu i, at man ved Stoppeur iagttog Tiden for Udstrømningen af en Række lige store Portioner — paa 50,2 cm^3 — af Vædsken. Af Tidsiagttagelserne og Portionernes Volumen beregnedes de pr. Sek. udstrømmende Rumfang V. Tallene for disse opførtes i en Tabel sammen med de tilsvarende gennemsnitlige Trykhøjder. Disse var beregnede en Gang for alle under den Forudsætning, at hver af de aftappede Portioner var 50,2 cm^3 og under Forudsætning af, at Centrum af Straalehullet laa 0,185 cm over Vædskebunden. Var den sidste Højde en anden, rettedes der for »Nulpunktsfejlen«. Derimod rettedes der ikke for de smaa svingende Afvigelser fra Værdien 50,2 cm^3 af de aftappede Portioner. Gennemsnitshøjderne beregnedes som det simple Middeltal af Højderne for Overfladen før og efter Udstrømningen af den paagældende Vædskeportion. Herved begaas en Fejl givet ved Udtrykket

$$\Delta h = \frac{h_1 + h_2 - 2\sqrt{h_1 h_2}}{4},$$

naar h_1 resp. h_2 er den effektive Begyndelses- resp. Slutningshøjde. Denne Fejl kunde praktisk set altid betragtes som forsvindende. Tabellen afbildedes endelig i en V^2-h-Kurve, d. v. s. en Kurve, der til Abscisse har Trykhøjden h, til Ordinat Kvadratet V^2 paa det pr. Sekund udstrømmende Vædskevolumen. Betegnes den til Overfladespændingens Modtryk svarende Vædskehøjde ved h_0, er den ved Torricellis Lov bestemte Værdi for V

$$V = \frac{\pi}{4} \varkappa^2 d^2 \cdot \sqrt{2g\,(h - h_0)},$$

hvor $\varkappa$ er Udstrømnings- eller Kontraktionskoefficienten (den effektive), og d Straalehullets Diameter. Ligningen for V^2-h-Kurven skulde derfor være

$$V^2 = \left(\frac{\pi}{4}\varkappa^2 d^2\right)^2 \cdot 2g \cdot (h - h_0).$$

Hvis altsaa $\varkappa$ er konstant, skulde V^2-h-Kurven vise sig at være en ret Linie, der af h-Aksen afskærer Stykket h_0, hvoraf Overfladespændingen C ifølge Christiansen skal kunne beregnes ved Udtrykket

$$h_0 = \frac{2C}{d'\rho g},$$

naar d' betegner Straalens Diameter og ρ Vædskens Vægtfylde.

A. Forsøg med Kvægsølv.

Ændringer i Straalens Udseende under Udstrømningen.

Efterhaanden som Overfladen synker i Reservoiret under Udstrømningsforsøget, bøjes Straalen mere og mere nedad af Tyngden. Samtidig bliver dens Overflade — saa længe Trykhøjden endnu ligger over en vis Grænse — mere og mere glat og regelmæssig. Henimod Slutningen af Udstrømningsforsøget antager Straalen et Udseende, der i væsentlig Grad afhænger af Straalehullets Diameter. Ved videre Boringer, for Eks. Boringer paa 2 mm og derover, vedbliver Straalen at være glat og danner, lige til den saa at sige har Hastigheden Nul, en sammenhængende, flere cm lang Straale. Fig. 2 a—e gengiver en Række Øjebliksfotografier af en saadan Straale, fremstillet ved et Straalehul i Karton paa 1,96 mm. Umiddelbart før Udstrømningen ophører, bliver ret pludselig den lange Straale instabil, og Udstrømningen finder nu Sted paa en ejendommelig pulserende Maade, hvorved Straalen, der nu er ganske kort, antager et flimrende Udseende. Ved noget snævrere Boringer — f. Eks. en Boring paa 1,2 mm — indtræder Instabiliteten ved en større Trykhøjde. Ved Boringer af en Diameter

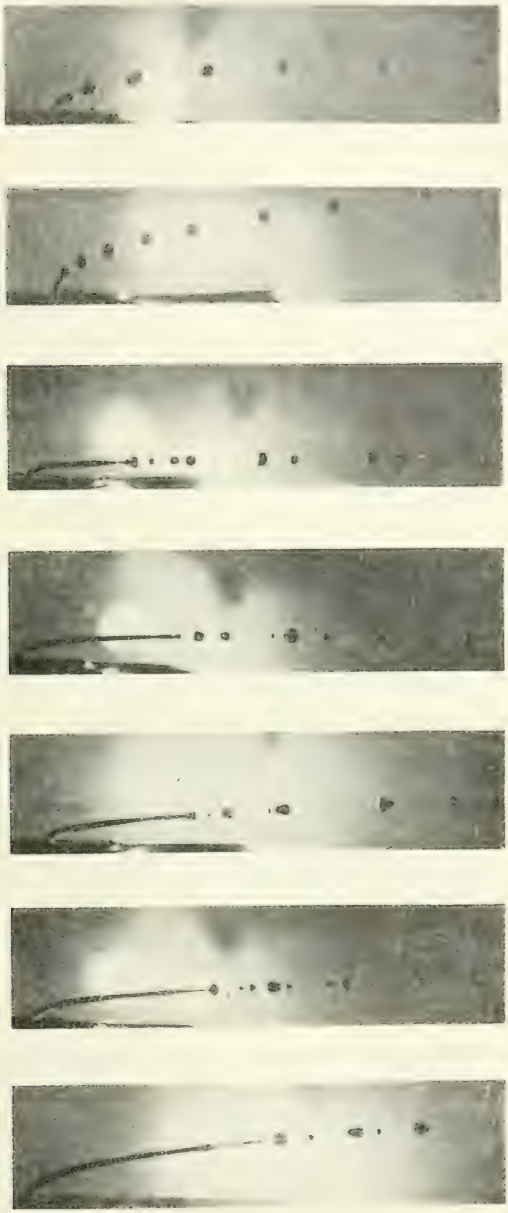


Fig. 2.

omkring 1,4 mm kan man iagttage det ejendommelige Fænomen, at Straalen slaar over fra den glatte Form til den pulserende og derefter atter til den glatte. Fig. 2 f og g viser to Øjebliksbilleder af en »Straale« paa 1,432 mm ved

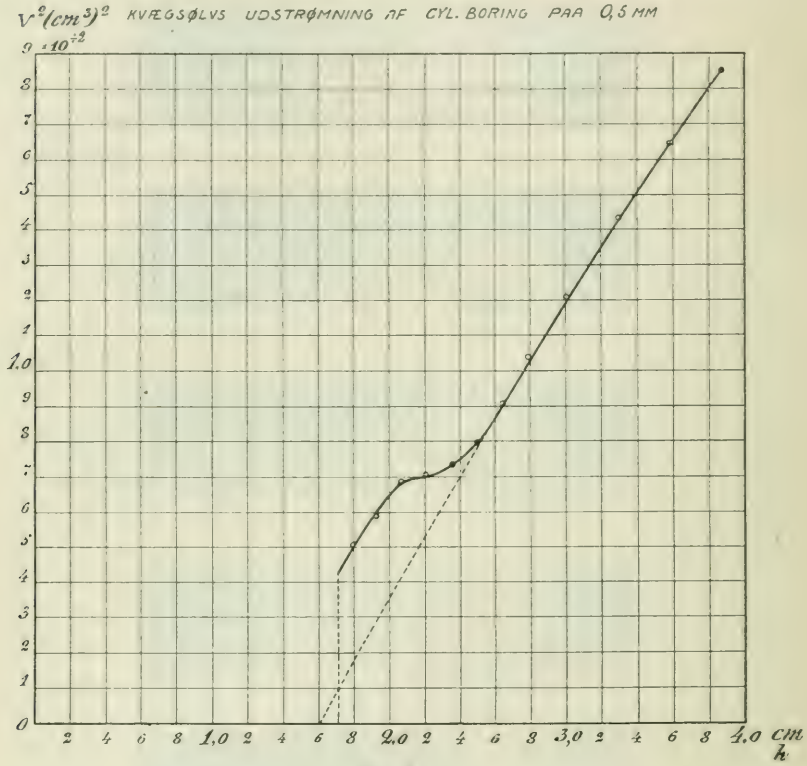


Fig. 3.

pulserende Udstrømning. Ved meget snævre Boringer paa Brøkdele af en Millimeter bliver den pulserende Udstrømningsform yderst udpræget, idet Kvægsølvet ender med i Draabeform at pible frem af Straalehullet.

Det fremgaar af de udførte Forsøg, at V^2-h-Kurven ved cylindriske Boringer er retliniet ned til de mindste Tryk.

Overgangen fra glat til pulserende Straale ændrer ikke Retningen væsentligt; men bliver Udstrømningen »piblende«, viser det sig yderst tydeligt i V^2-h-Kurven, saaledes som det ses af Fig. 3. Da det gennemsnitlige fra Overfladespændingen hidrørende Tryk inde i de frempiblende Kvægsølvdraaber er mindre end det hydrostatiske Tryk hidrørende fra Overfladen af en Kvægsølvcylinder af Boringens Diameter, strømmer der mere Kvægsølv ud, end der vilde ved en Straale. Derfor krummer V^2-h-Kurven, naar Draabedannelsen begynder, først med Hulheden opad for senere atter at falde jævnt af som Følge af det formindskede Udstrømningstryk.

Udstrømningens Ophør.

Naar Trykket er naaet ned til en vis Grænse, hører Udstrømningen pludselig op, og Kvægsølvet stiller sig i Straalehullet med en krum Overflade, saaledes som det er antydet i Fig. 4. Spændingen i denne Overflade bærer nu den resterende Kvægsølvsøjle h_0'. Jeg foretog Bestemmelser af h_0' for en Række Boringer, dels koniske dels cylindriske, og jeg fandt, at Sikkerheden paa Maalingen af h_0' var meget betydelig. Eksempelvis anføres følgende Gentagelsesværdier for de sidst aftappede Vædskevolumina ved fire Udstrømningsforsøg.

Fig. 4.

Nr.	0,80. 90°	0,93. 90°	2,0. 90°	1,5 cyl.
1	9,2 cm³	43,5 cm³	1,7 cm³	7,8 cm³
2	8,5 -	40,1 -	1,7 -	6,7 -
3	8,2 -	42,5 -	1,9 -	6,7 -
4	8,8 -	45,0 -	1,6 -	
5	....	....	2,1 -	
	8,7 cm³	42,8 cm³	1,8 cm³	7,1 cm³
	$\Delta = 0{,}5$ cm³	$\Delta = 2{,}7$ cm³	$\Delta = 0{,}2$ cm³	$\Delta = 0{,}7$ cm³

Gentagelsesværdierne fandtes ved efter Afmaalingen af Restvoluminet atter at hælde dette tilbage i Beholderen og igen bestemme det Rumfang, som løb ud. Usikkerheden er, som det synes, abnorm stor i andet Tilfælde, og her

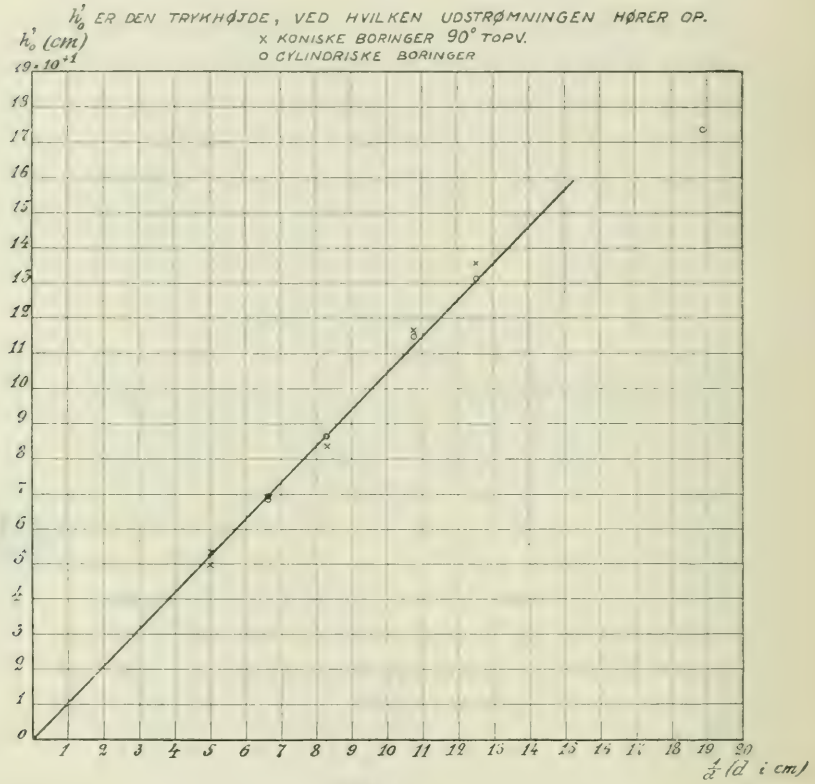


Fig. 5.

ligger den dog under 3 cm^3. Det vil sige, den Trykhøjde, ved hvilken Udstrømningen ophører, er ved een Iagttagelse bestemt med en Sikkerhed af ca. 0,16 mm. I de andre Tilfælde er Sikkerheden væsentlig større.

Dersom Vædskeoverfladen i Boringen Fig. 4 er kugleformet og har Krumningsradius r', skal r' være givet ved

$$\frac{2C}{r'} = h_0' \cdot 13{,}55 \cdot 981,$$

idet C betegner Kvægsølvets Overfladespænding.

Hvis det tør antages, at r' er proportional med Boringens Radius r, saa at r' kan sættes lig $k \cdot r$, skal man ved at afbilde Variationen af h_0' med $1/r$ eller $1/d$, hvor r og d er Radius resp. Diameteren i Straalehullet, finde en ret Linie. I Fig. 5 er alle under Arbejdet iagttagne Værdier for h_0' indtegnede i deres Afhængighed af $^1/_d$. Naar man ser bort fra Værdien svarende til Boringen paa 0,53 mm, der ogsaa ved de ovenfor betragtede Udstrømningsforhold udviste Anomalier, fordeler alle Forsøgspunkter sig med stor Regelmæssighed om en ret Linie gennem Begyndelsespunktet. Af den rette Linie finder man for Konstanten k Værdien 1,496, idet der for C er benyttet Værdien 520 $^{dyn}/_{cm}$. Kvægsølvet stiller sig altsaa ind med en Krumningsradius, der med god Tilnærmelse er lig 1,50 Gange Boringens Radius. Og Krumningsradius er ganske uafhængig af, om Boringens Profil er cylindrisk eller konisk.

Den extrapolerede Nulhøjde og dens Sammenhæng med Overfladespændingen.

Efter den Christiansenske Antagelse skal, som anført, det Stykke h_0 — den extrapolerede Nulhøjde —, som den forlængede V^2-h-Kurve afskærer af Abscisseaksen, være bestemt ved

$$\rho g h_0 = \frac{2C}{d'},$$

hvorved det erindres, at d' betegner Straalens Diameter. Det Resultat, mine Forsøg førte til, var i Korthed følgende. Ved cylindriske Boringer er V^2-h-Kurven retliniet. Forlænges den retliniet ned til Abscisseaksen, afskærer den et

Stykke (h_0), Fig. 6, der, indsat i Formlen ovenfor, altid giver samme Værdi for C, naar man for d' sætter Straalehullets Diameter d — svarende til, at den virkelige Kontraktion ved cylindriske Boringer konstant er Nul. Ved

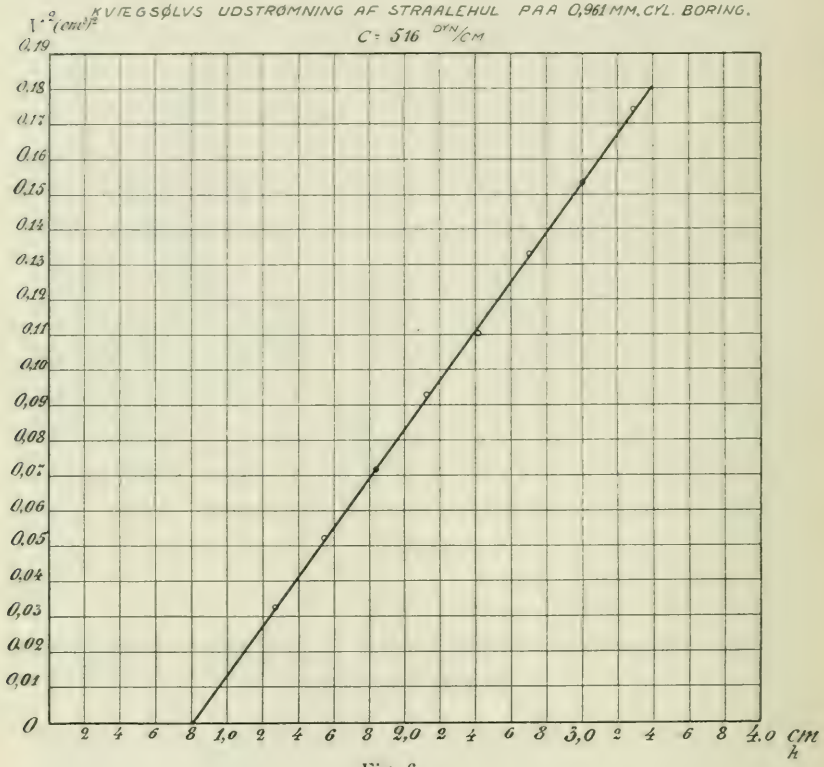


Fig. 6.

koniske Boringer er V^2-h-Kurven krum, Fig. 7. Fortsættes den ned til Abscisseaksen ved Hjælp af en bøjelig Lineal, der lægges an langs den experimentalt bestemte Del, og indsætter man i Formlen ovenfor den afskaarne Del for h_0, samt sætter man for d' ligesom ved de cylindriske Boringer Straalehullets Diameter — svarende til, at den virkelige Kontraktion ved aftagende Hastighed tenderer mod

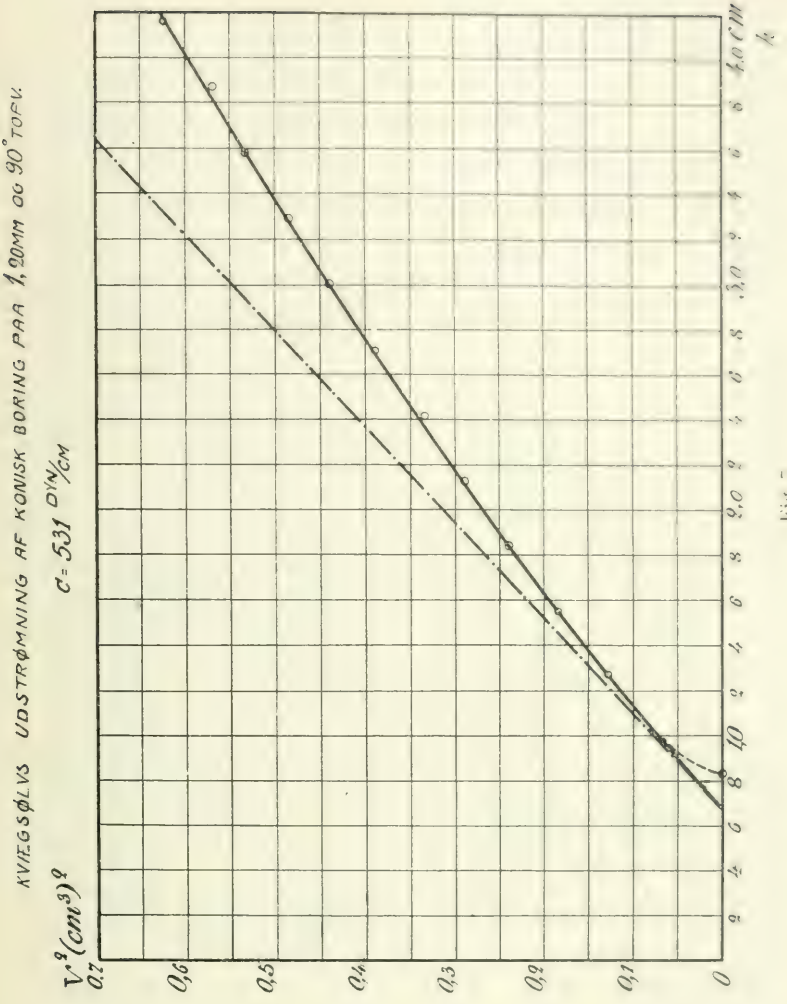


Fig. 7

Nul[1] —, faar man atter for C en konstant Værdi og i Hovedsagen den samme, der blev fundet ved de cylindriske Boringer. Den fælles C-Værdi falder herved meget nær

[1] Angivelserne om Kontraktionen gengiver Resultater af en Række Undersøgelser, der vil blive beskrevet andetsteds.

sammen med de Bestemmelser af Kvægsolvs Overfladespænding, der maa anses for de hidtil bedste.

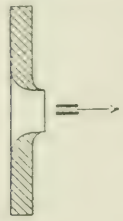


Fig. 7.

Til Belysning af, hvor nøje V^2-h-Kurven baade ved Straaler uden og med Kontraktion gaar igennem det ved Overfladespændingen C bestemte Punkt af Abscisseaksen anføres en Række Værdier for C, fundne ved det ovenfor antydede Forsøg paa Uddannelse af en Metode til Bestemmelse af Overfladespænding.

Tabel I.

Nr.	Huldiameter	Længde af Boringens cylindriske Del	C	Anm.
5	0,799 mm	2,0 mm	506 dyn/cm	Cylindriske Huller borede i Staal. C-Værdierne er korrigerede for Tryktabet i Boringernes cylindriske Del. (Se nedenfor).
4	0,961 -	1,2 -	516 —	
6	1,218 -	3,0 -	527 —	
9 og 10	1,507 -	2,5 -	507 —	

Middeltal 514 dyn/cm ved ca. 15°

Tabel II.

Nr.	Huldiameter	Kartonens Tykkelse	C	Anm.
85	1,228 mm	0,25 mm	530 dyn/cm	Huller pressede i Karton.
91 A og B	1,923 -	0,13 -	547 —	Straalen udviser foranderlig Kontraktion.
93	1,228 -	0,25 -	548 —	
94	0,965 -	0,13 -	532 —	
95	0,952 -	0,13 -	540 —	
96	1,380 -	0,13 -	522 —	

Middeltal 536 dyn/cm ved ca. 16°

Som det fremgaar af Forsøgsnumrene, gengiver Tallene i første Tabel Resultater fra nogle af de allerførste Bestemmelser. Disse Bestemmelser maa betragtes som orienterende,

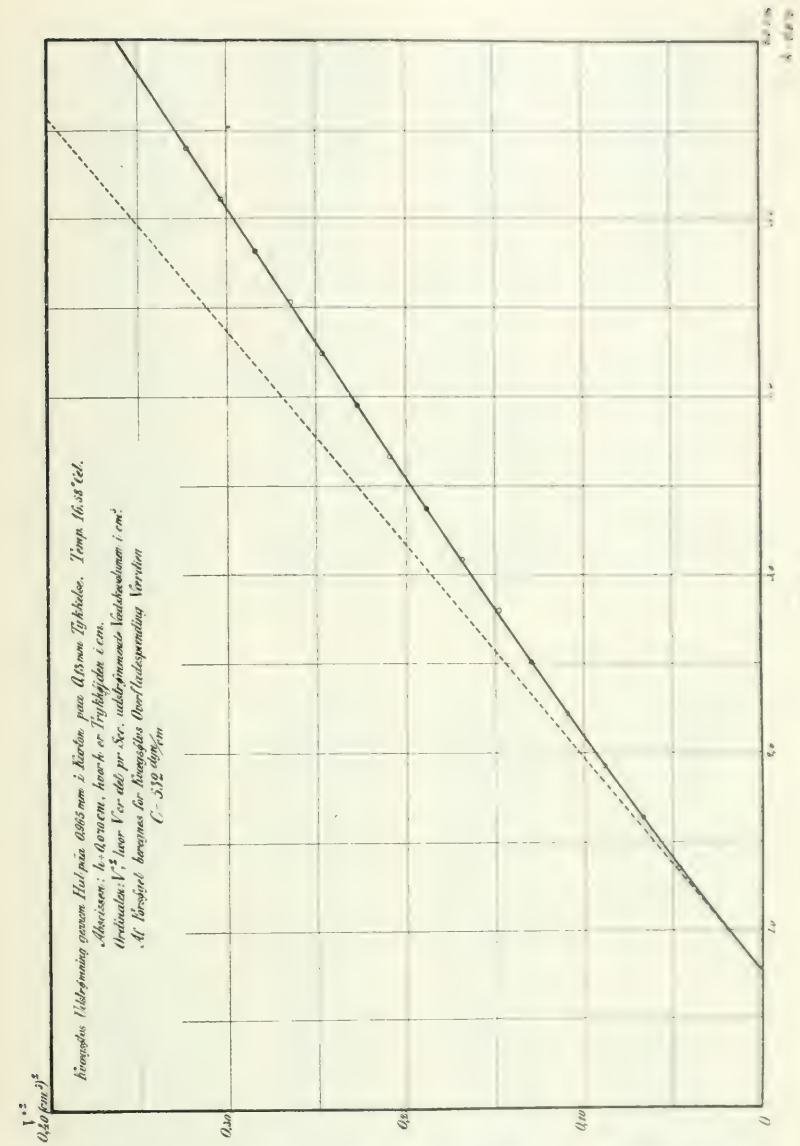


Fig. 9

anden Tabel derimod indeholder de sidste og endelige Bestemmelser i den hele Forsøgsrække. Den sande Værdi for Overfladespændingen ved 20° af en ny Kvægsølvoverflade synes at

ligge omkring 540 $^{dyn}/_{cm}$. Til Belysning af den Punktsikkerhed, som opnaaedes ved de senere Udstrømningsforsøg, der alle udførtes med Huller pressede i Karton og med den i Fig. 8 viste Profil gengives i Fig. 9 V^2-h-Kurve fra Forsøg Nr. 94.

B. Forsøg med Vand.

Forsøgsbetingelser.
Straalens Ændringer under Udstrømningsforsøget.
Udstrømningens Ophør.

At udføre et entydigt Udstrømningsforsøg ved lavt Tryk med et Straalehul i Metal eller Glas lader sig ikke gøre med Vand eller nogen Vædske, der væder det paagældende Materiale. Der samler sig straks om Straalehullet en Vædskemasse, saaledes som det er antydet i Fig. 10. Ved Trykhøjder paa 3—4 cm presses Straalen frem af denne Masse og staar med en veldefineret og glat Overflade, men den pr. Sek. udstrømmende Vædskemængde kan variere mange Procent som Følge af tilfældige Ændringer i Vædskeansamlingens Form, og den kan være flere Procent større end den Mængde, der vilde strømme ud ved samme Trykhøjde, saafremt Straalen stod med skarp Kant imod Hulpladen. Ved noget lavere, men dog ved et relativt højt Tryk ophører Udstrømningen i Straaleform ganske, idet Vandet driver ned langs Hulpladens lodrette Overflade. Vandets Vedhængning ved Hulpladen umuliggør derved praktisk set Optagelsen af den nederste Del af en normal V^2-h-Kurve og som Følge heraf ogsaa Extrapolation til Nulhøjden h_0.

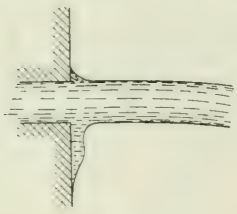


Fig. 10.

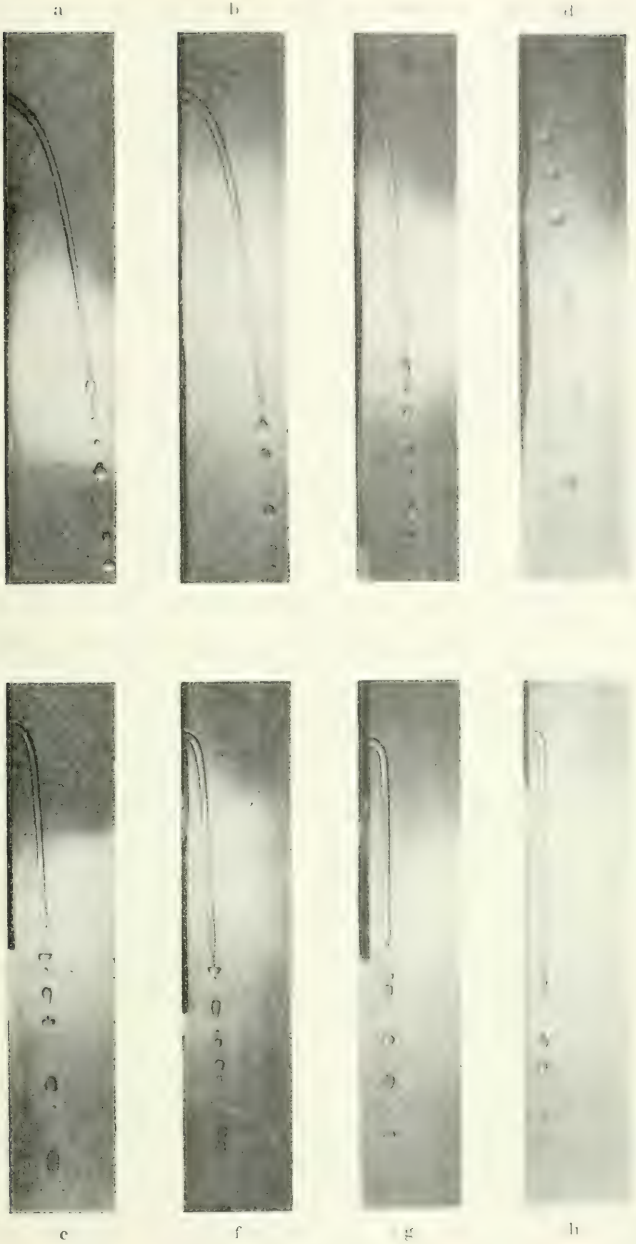


Fig. 11.

Men man kan opnaa meget nær lige saa sikre Forsøgsbetingelser ved Vand som ved Kvægsølv, naar man ved den første Vædske til Fremstilling af Straalehulspladen benytter Parafin eller bedre parafineret Papir eller Karton. Overfor Parafin opfører Vand sig i Virkeligheden ganske paa samme Maade som Kvægsølv overfor Jærn. Et fuldstændigt Udstrømningsforsøg er derfor muligt, og man iagttager under et saadant ganske de samme Ændringer i Vandstraalens Udseende som ovenfor er skildrede for Kvægsølvstraalens Vedkommende. Særlig udpræget er den pludselige Overgang ved lave Tryk fra en lang glat Straale til en pulserende Udstrømning, der saa atter ved et endnu lavere Tryk slaar om til en glat Straale. Fig. 11 a—h gengiver i Øjebliksfotografier de tre paa hinanden følgende Faser i Udstrømningen. Ret karakteristisk er den bratte Maade, hvorpaa den sammenhængende Del af Straalen hører op ved pulserende Udstrømning. Naar Trykhøjden er naaet ned under en vis Grænse, ophører Udstrømningen i Straaleform, og Vandet rinder nu ned langs Hulpladens Væg. Ved et endnu lavere Tryk h_0' hører Udstrømningen ret pludseligt op, og Vandet stiller sig ind i Boringen med en kuplet Overflade — ganske som Tilfældet var ved Kvægsølv. Den Sikkerhed, hvormed h_0' kan bestemmes ved Vand, er dog ikke saa stor som ved Kvægsølv. Saaledes fandtes ved gentagen Bestemmelse af det sidst udstrømmende Vædskevolumen (Restvoluminet) følgende Værdier

$$29{,}4 \text{ — } 24{,}4 \text{ — } 38{,}4 \text{ — } 40{,}8 \text{ cm}^3,$$

i hvilken Række en ensidig Ændring vistnok gør sig gældende. Ved fire Forsøg fandtes for h_0' de i hosstaaende Tabel anførte Værdier.

Forsøg	Boringens Radius r	h_0'	r'	r'/r
Nr. 61	0,735 mm	1,318 cm	1,137 mm	1.55
- 74	0,885 -	1,051 -	1,427 -	1.61
- 63	0,901 -	0,945 -	1,590 -	1.76
- 73	1,091 -	0,713 -	2,064 -	1.89

I Tabellen betegner r' Radius i Kuplen beregnet af Udtrykket

$$\frac{2C}{r'} = h_0' g,$$

hvor C — Vandets Overfladespænding — er sat lig 73,6 $^{dyn}/_{cm}$.

Forholdet mellem r' og Straalehullets Radius r vokser, som det ses, med Straalehullets Radius. Jo større Hullet er, desto fladere synes Kuplen altsaa at blive. Forsøg Nr. 61 og 63 blev udført med cylindriske Boringer i Parafinpropper, Nr. 73 og 74 med Huller pressede i parafineret Papir.

Nulhøjden og Overfladespændingen.

Resultatet af min Prøve af Christiansens Antagelse var for Vand ganske det samme som for Kvægsølv. Den extrapolerede Nulhøjde h_0 bestemmer Vandets Overfladespænding, naar den indsættes i det ovenfor anførte Udtryk, og naar d i dette Udtryk sættes lig Straalehullets Diameter. Den Sikkerhed, hvormed man kan finde Overfladespændingen, vil fremgaa af følgende Oversigt over en Del af de udførte Bestemmelser.

Forsøgene i Tabel I er nærmest at betragte som orienterende. Ved Forsøgene i Tabel II blev Fejlkilderne undersøgt med Omhu, og hvad Tabel III angaar, repræsenterer de deri anførte Tal de sidste og formentlig bedste af alle udførte Bestemmelser. Som det ses, ligger alle fundne

Tabel I.

Forsøg	*d*	*C* destilleret Vand	*C* Vand fra Vandhanen	Anm.
Nr. 32	1,019 mm	74,8 dyn/cm	— dyn/cm	Korte cylindriske
- 30	1,220 -	74,4 —	— —	Boringer i Para-
- 33 (26)	1,591 -	75,2 —	75,3 —	fin med konisk
- 34 (25)	1,996 -	70,0 —	70,0 —	Indledning.
- 31	2,211 -	73,0 —	— —	

Middeltal 73,5 dyn/cm ved ca. 10°

Tabel II.

Forsøg	Boringens Længde	*d*	*C* ukorrigeret	*C* korrigeret	Anm.
Nr. 57	0,5 cm	1,616 mm	70,7 dyn/cm	70,2 dyn/cm	Rent cylindriske
- 58	0,2 -	1,567 -	77,1 —	70,1 —	Boringer i Pa-
- 59	0,1 -	1,698 -	80,3 —	72,3 —	rafin. Indre
- 60	0,4 -	1,442 -	88,5 —	72,9 —	Kant afrundet.
- 61	0,8 -	1,470 -	101,3 —	72,6 —	
- 62	0,8 -	2,490 -	96,0 —	73,9 —	
- 63	0,8 -	1,801 -	90,7 —	70,5 —	

Middeltal 71,8 dyn/cm

Korrigeret for forskellige Fejl 74,8 dyn/cm ved ca. 10°

Tabel III.

Forsøg	Hulpladens Tykkelse	*d*	*C*	Anm.
Nr. 108	0,17 mm	1,628 mm	72,8 dyn/cm	Huller pressede i
- 118	0,17 -	1,624 -	74,8 —	parafineret Papir.
- 111	0,17 -	1,878 -	74,4 —	Indre Rand af-
- 109	0,38 -	1,868 -	74,1 —	rundet.

Middeltal 74,0 dyn/cm ved 10°

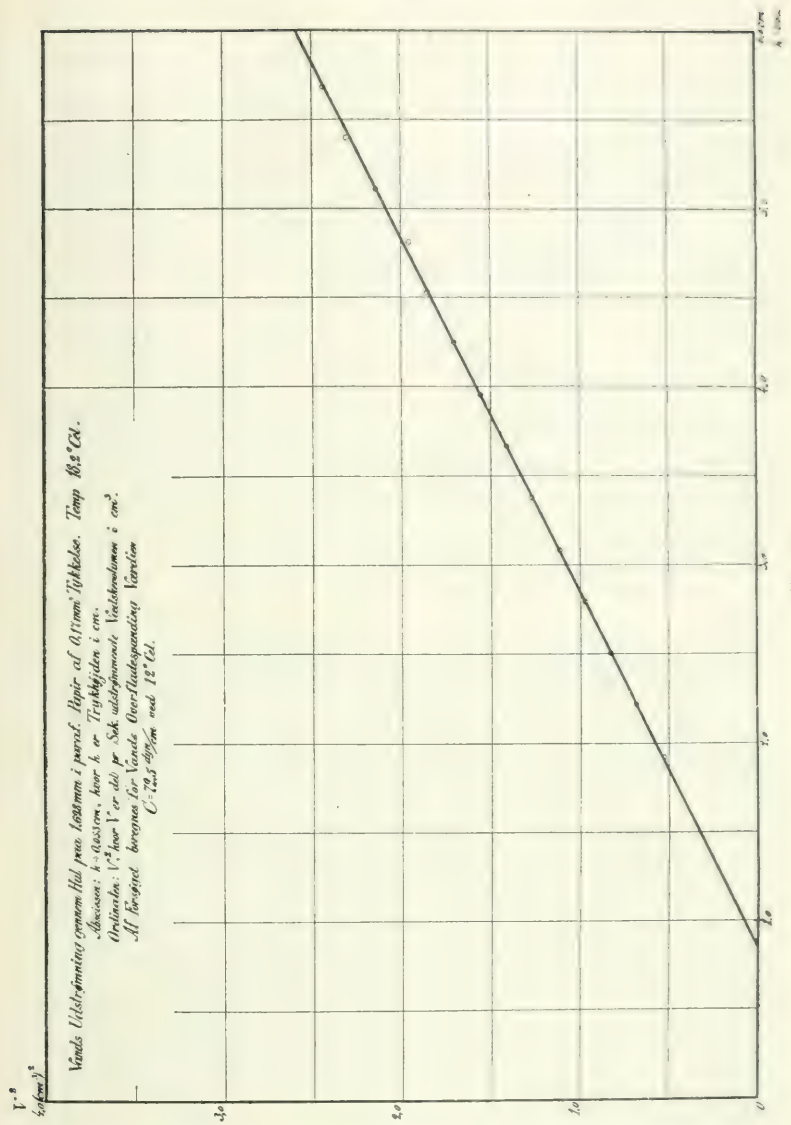


Fig. 12.

C-Værdier meget tæt ved 74,0 dyn/cm eller ved den Værdi, man sædvanligvis betragter som den rigtige for Vands Overfladespænding ved 10°. Andre Forsøgsrækker blev gennem-

fort, der i Hovedsagen gav samme Resultat. Rækken Nr. 49—55, der udførtes med cylindriske Boringer med konisk Indledning, gav for C Værdien 72,4 $^{dyn}/_{cm}$ ved 10°. Rækken Nr. 71—74b (8 Bestemmelser ved stærkt varierede Temperaturer) gav $C = 73{,}4$ $^{dyn}/_{cm}$ ved 10°. Endelig gav det isolerede Forsøg Nr. 69 ved krumliniet Extrapolation Værdien 75,2 $^{dyn}/_{cm}$ ved 10°.[1] For at give en Forestilling om, hvor sikkert V^2-h-Kurven ved disse Forsøg kunde bestemmes — og Extrapolationen udføres — gengives i Fig. 12 Forsøg Nr. 108.

Vand gav altsaa, hvad Spørgsmaalet om Sammenhængen mellem h_0 og C angaar, ganske samme Resultat som Kvægsølv. Overfladespændingens Modtryk maales direkte ved h_0.

Gnidningstabet i en længere cylindrisk Boring.

I Tabel II ovenfor er der i Overskriften over sidste Talkolonne antydet en Korrektion, som nu skal omtales noget nærmere. I den foregaaende Kolonne staar de Værdier for C, der udmaaltes direkte paa V^2-h-Kurverne. Som det ses, er disse Værdier gennemgaaende for store, og desto mere for store, jo længere og snævrere den cylindriske Boring var. Det gik op for mig, at Aarsagen maatte søges i Gnidningen i Boringen, og jeg forsøgte nu at rette Fejlen ud fra den Betragtning, at Gnidningen i Boringen betød et Tryktab h_p, der maaske lod sig beregne ved Poiseuilles Udtryk, altsaa ved

[1] Huller pressede i parafineret Karton gav ved Vand i Almindelighed en retliniet V^2-h-Kurve. I Forsøg Nr. 69 var Kartonen temmelig tynd, og Hullets indvendige Kant har formentlig været ret skarp, saaledes at en virkelig Kontraktion har kunnet gøre sig gældende.

$$h_p = \frac{128 \cdot \eta \cdot L}{\pi d^4 g} \cdot V,$$

hvor η er Koefficienten for Vands indre Gnidning ved Forsøgstemperaturen, L Boringens Længde, og V det pr. Sek. udstrømmende Vædskevolumen.

Voluminet V skulde herefter være bestemt ved

$$V = \frac{\pi}{4} \varkappa^2 d^2 \cdot \sqrt{2g(h - h_0 - h_p)}$$

i Stedet for ved

$$V = \frac{\pi}{4} \varkappa^2 d^2 \cdot \sqrt{2g(h - h_0)}.$$

Jeg reducerede nu den faktisk fundne V^2-h-Kurve til den, der skulde være kommen, saafremt h_p ikke gjorde sig gældende. Reduktionen foretoges som antydet i Fig. 13 ved for to eller tre Punkter at beregne Værdierne for h_p og formindske Punkternes Abscisser med den beregnede h_p-Værdi. Den oprindelige V^2-h-Kurve, der var meget nær retliniet eller maaske havde en Antydning af en Krumning med Hulheden opad, blev herved forvandlet til en ny, meget nær retliniet, men lidt stejlere, V^2-h-Kurve, der forlænget retliniet afskar et Stykke h_0 af Abscisseaksen givet, som det fremgaar af Tabel II, ved

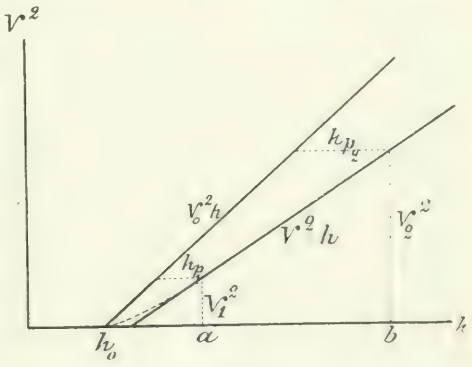


Fig. 13.

$$\frac{2C}{d} = h_0 g.$$

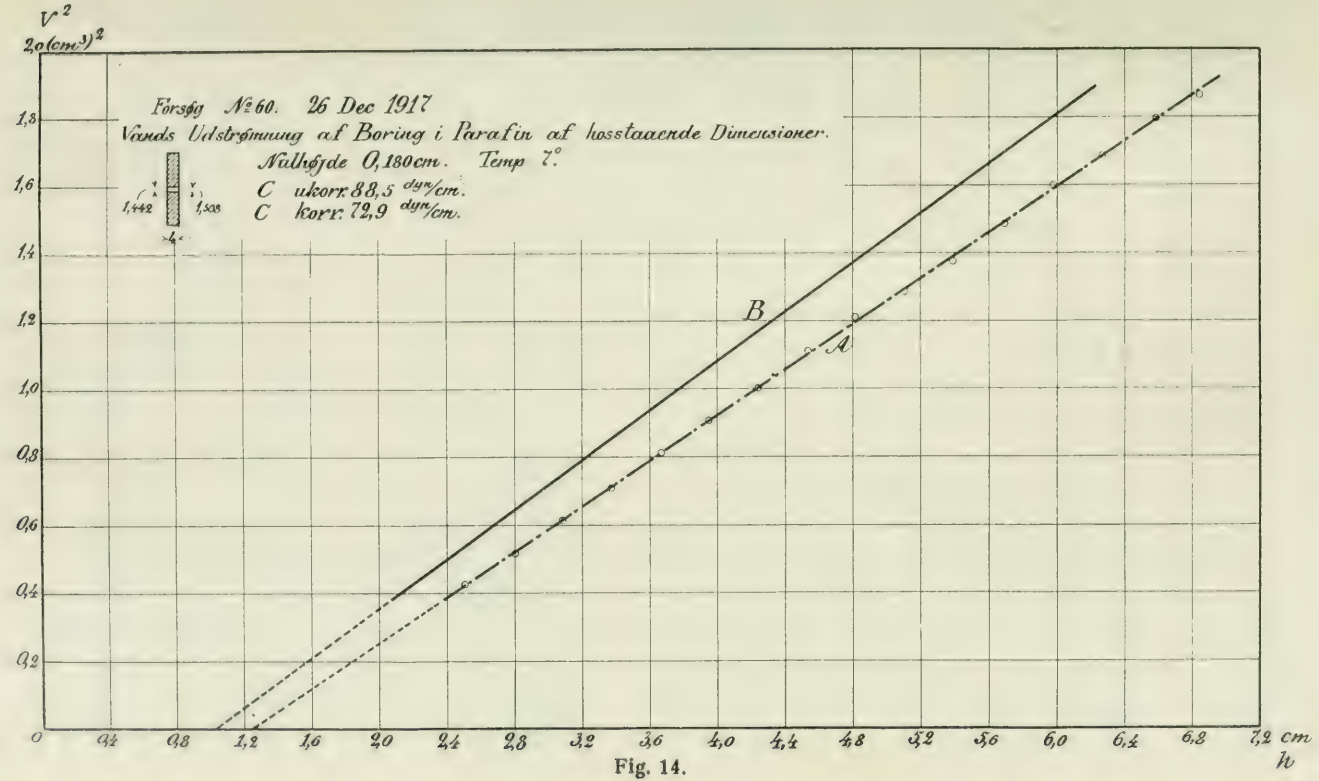

V^2
$20(cm^3)^2$
1,8
1,6
1,4
1,2
1,0
0,8
0,6
0,4
0,2
0
0,4
0,8
1,2
1,6
2,0
2,4
2,8
3,2
3,6
4,0
4,4
4,8
5,2
5,6
6,0
6,4
6,8
7,2 cm
h
Forsøg № 60. 26 Dec 1917
Vands Udstrømning af Boring i Parafin af lodstaaende Dimensioner.
Nålhøjde 0,180 cm. Temp 7°.
C ukorr. 88,5 dyn/cm.
C korr. 72,9 dyn/cm.
1,442
1,508
B
A

Fig. 14.

Ved Beregningen af h_p benyttedes for η den Værdi for den indre Gnidning, der svarer til en Parallelstrømning. For at give en Forestilling om et Forsøg, behandlet paa den her antydede Maade, gengives i Fig. 14 Bestemmelsen Nr. 60.

Ætherdamps Indflydelse paa Vands Overfladespænding.

Jeg fandt det af Interesse at gentage det gamle kendte Forsøg over Indflydelsen af Ætherdamp paa Vands Overfladespænding. Ved Forsøget benyttedes en kort cylindrisk Boring paa 1,278 mm i massiv Parafin. Forsøget udførtes paa den Maade, at Udstrømningstiden for hver af de normale Portioner paa 50,2 cm^3 iagttoges én Gang, idet Straalen traadte ud i Luft, og én Gang, idet der over Straalen tæt op til denne anbragtes en med Æther vædet Vattot. Fig. 15 gengiver Resultatet af Dobbeltforsøget. Det bør bemærkes, at V^2-h-Kurven ved Ætherdamp er nøjagtig parallel med Kurven ved ren Luft. Dette tyder paa, at Virkningen af Ætherdampen gør sig gældende næsten momentant, thi ellers kunde det vel næppe være ligegyldigt, om Straalen har en Hastighed paa 30 cm, — som den har det ved de nederste Forsøgspunkter, — eller 60 cm — som ved de øverste Punkter. Man vil iøvrigt lægge Mærke til, at Punktusikkerheden ved Ætherdamp ikke er større end ved ren Luft. Dette gør det sandsynligt, at det er den maximale Virkning af Ætherdamp, man har med at gøre, d. v. s. den, der vilde iagttages ved Udstrømning i en Atmosfære alene bestaaende af Ætherdamp. Forandringen i Overfladespænding er, som det ses, 18 %. Den gør sig naturligvis tydeligst gældende, naar Straalen har sin mindste Hastighed. Her er Virkningen yderst paafaldende og egner sig til Demonstration. Naar Straalen i Luft antager den tidligere

beskrevne urolige eller pulserende Karakter, ser man, at den retter sig op og samler sig til en fuldstændig glat og sammenhængende Straale. — Jeg forsøgte, om en kendelig Virkning kunde iagttages ved andre flygtige Stoffer (Ben-

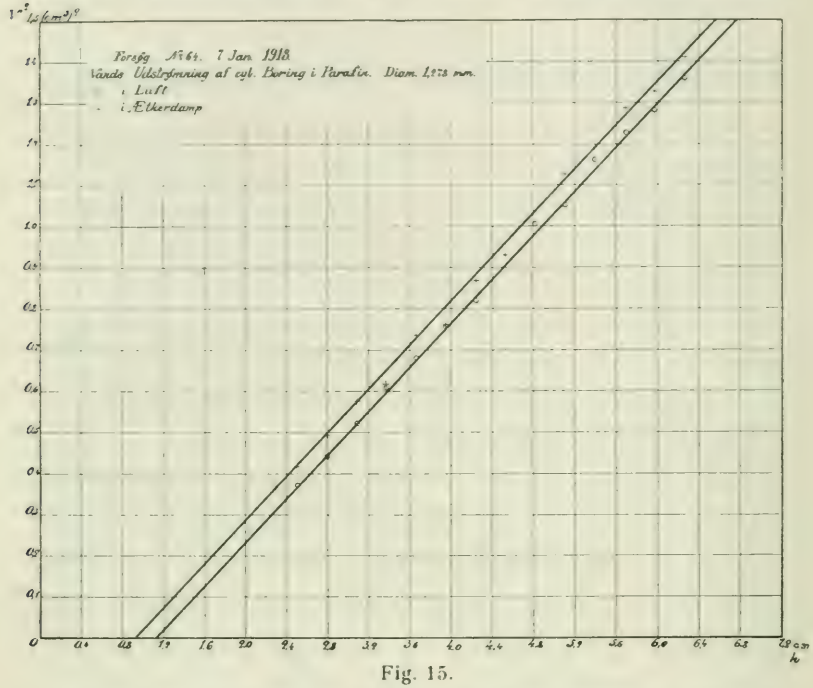


Fig. 15.

zin, Acetone, Hexan, konc. Svovlsyre, Ammoniak og Kamfer), men Resultaterne var altid negative eller i ethvert Fald uafgørende.

Nogle Iagttagelser og deres Fortolkning.

De ovenfor omtalte Forsøg har vist, at Overfladespændingen C under Udstrømning af en Vædske gennem et cirkulært Hul gør sig gældende som et Modtryk lig med det hydrostatiske Tryk $\frac{2C}{d}$ indeni en Vædskecylinder af Straalehullets Diameter d. Ved koniske Boringer gælder

dette dog altsaa kun ved de aller laveste Hastigheder, fordi Straalen først ved disse Hastigheder antager cylindrisk Form. Ved cylindriske Boringer gælder det for alle Hastigheder, svarende til at Straalens virkelige Kontraktion permanent er Nul. Vigtigt synes det mig nu at lægge Mærke til, at det nævnte Modtryk i begge Tilfælde gør sig gældende ved de laveste Tryk. Ved disse Tryk er den største Del af Straalen nemlig meget langt fra at være cylindrisk. I Virkeligheden er kun den aller inderste Del det, f. Eks. den første hele eller halve Millimeter. Længere behøver den cylindriske Del af Straalen altsaa ikke at være, for at der i Straalens første Del skal indstille sig et Modtryk bestemt ved $\frac{2C}{d}$.

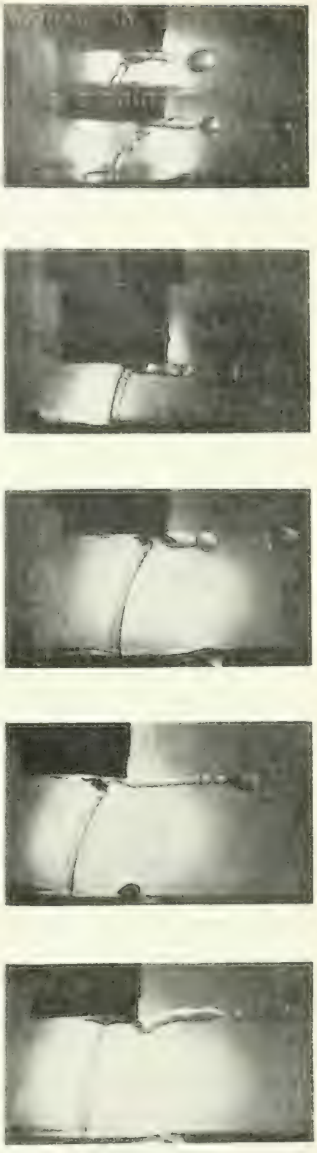


Fig. 16.

Jeg udførte et Forsøg, der yderligere belyser, hvor uafhængig det udstrømmende Volumen, altsaa det hydrostatiske Modtryk umiddelbart foran Straalehullet, er af Straalens Form blot én eller et Par Millimeter fra Straalehullet. Ved dette Forsøg ramte Straalen en Plade, og der dannedes da

paa Straalen de kendte Ringbølger. En Række Øjebliksbilleder af disse ses i Fig. 16 a—e. Det viste sig, at Tiden for Udstrømningen af et bestemt Volumen er uafhængig af Pladens Afstand fra Straalehullet, naar Afstanden blot ikke er saa kort, at de første faa Tiendedele Millimeter af Straalen ændrer Form. Selv ved en saa deformeret Straale som den sidste af de i Fig. 16 viste, var der næppe nogen Indflydelse af Deformationen at spore.

Af de anførte Iagttagelser drager jeg den Slutning, at hvis blot Straalen er cylindrisk paa nogle faa Tiendedele Millimeter, bestaar der under denne cylindriske Overflade et Tryk bestemt ved $\frac{2C}{d}$ (i ethvert Fald, naar d er af en Størrelsesorden paa 1—2 mm som ved de her betragtede Forsøg). Mere almindeligt drager jeg den Slutning, at det hydrostatiske Tryk inde i Straalen tro afspejler Formvariationer i Straalens Overflade, selv om disse Variationer er meget bratte. Ud fra dette Synspunkt er jeg i Stand til at forklare mig en Række andre Iagttagelser, i Særdeleshed følgende to. Hvis en iøvrigt cylindrisk Vandstraale lige netop væder Kanten af Straalehullet, danner Overfladen langs Kanten en dobbelt krummet, ringformet Flade, der vender Hulheden udad. Den hule Del af Fladen er undertiden næppe til at se, strækker sig f. Eks. kun over de første Par Tiendedele Millimeter af Straalen, og dog bevirker den straks, at det udstrømmende Volumen sættes meget kendeligt op. Grunden hertil er efter det anførte den, at den første Del af Straalen med den dobbelte Krumning, hvor ringe i Udstrækning den end er, i det væsentlige bestemmer det hydrostatiske Tryk i Mundingen af Straalehullet og derved det udstrømmende Vædskevolumen. Det hydrostatiske Tryk under den dobbelt krummede Overflade er givet ved

$$2C\left(\frac{1}{d_1}+\frac{1}{d_2}\right),$$

hvor d_1 og d_2 er Hovedkrumningsdiametrene i Overfladen. Af disse er den ene f. Eks. d_2 negativ, medens den anden er af samme Størrelsesorden som Straalehullets Diameter d, omend noget større. Modtrykket maa altsaa blive mindre end ved den rent cylindriske Straale.

Den anden Iagttagelse angaar Formen af V^2-h-Kurven ved koniske Boringer. Dersom man i Fig. 17, der fremstiller en saadan Kurve, betragter en begrænset Del c' b', ses det, at Udstrømningen indenfor dette Omraade finder Sted med en effektiv Kontraktion svarende til Hældningen af Stykket c' b' og med et Modtryk givet ved h_0'', der vil være mindre end h_0, og som i visse Tilfælde vistnok endog kan blive negativ. Den lavere Værdi for h_0 maa ganske sikkert igen forklares ved, at Straalen, der ved den koniske Boring og det højere Tryk udviser en virkelig Kontraktion, tæt udenfor Straalehullet har en dobbelt Krumning med én negativ Krumningsradius. Ændringen med Trykhøjden af det Stykke, Tangenten til V^2-h-Kurven afskærer af h-Aksen, afspejler altsaa efter denne Forklaring, den jævne Ændring i den virkelige Kontraktion fra den konstante Værdi, denne har ved store Hastigheder, til Værdien Nul, som den har ved Hastigheden Nul.

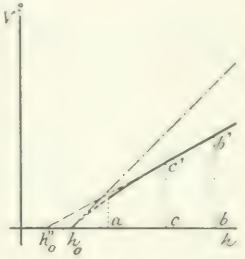


Fig. 17.

Den virkelige V^2-h-Kurve ved en cylindrisk Boring med kendelig Gnidning krummer, som Fig. 13 skal antyde, svagt med Hulheden opad. Krumningen er ved Straalehuller som dem, vi ovenfor har betragtet, kun tydelig paa Kurvens nederste Del, for hvilken der ikke kan fremskaffes direkte Iagt-

tagelser. Det vil fremgaa af følgende Udledning. Vi har set, at en Formindskelse af alle Abscisser med de ved Poiseuilles Formel

$$h_p = \frac{128 \cdot \eta \cdot L}{\pi g d^4} \cdot V$$

beregnede Beløb reducerer den virkelige V^2-h-Kurve, der tilsyneladende sigter mod h_0'' af Abscisseaksen til en V^2-h-Kurve, der sigter mod h_0. Indføres h_p i Torricellis Udtryk for Udstrømningshastigheden, saaledes at dette bliver

$$v = \varkappa_f^2 \sqrt{2g(h - h_0 - h_p)},$$

faas for V^2-h-Kurven Ligningen:

$$V^2 - 16\pi\varkappa^4\eta L \cdot \sqrt{V^2} = \left(\frac{\pi}{4} d^2\right)^2 \cdot \varkappa^4 \cdot 2g(h - h_0)$$

i Stedet for som ved den tidligere meddelte simplere Teori

$$V_0^2 = \left(\frac{\pi}{4} d^2\right)^2 \cdot \varkappa^4 \cdot 2g(h - h_0).$$

Ved Korrektionen reduceres den virkelige V^2-h-Kurve til en Kurve med den sidste Ligning.

Betragter vi nu de to sidste Udtryk, ses det, at Indflydelsen af Tryktabet er den, der er søgt anskueliggjort i Fig. 13. Den faktiske og den reducerede Kurve gaar begge gennem Punktet $h_0, 0$, men medens den reducerede V_0^2-h-Kurve skærer Abscisseaksen i dette Punkt, tangerer den faktiske V^2-h-Kurve Aksen. Den sidste Kurve vil krumme som antydet, men Krumningen vil snart blive meget ringe. Hvorledes Talforholdene iøvrigt kan stille sig, vil fremgaa af efterfølgende Tabel over $\frac{V_0^2 - V^2}{V^2}$ og $\frac{d^2V^2}{dh^2} \Big/ \frac{dV^2}{dh}$. Ved Tabellens Beregning er $\varkappa$ sat lig 1 og η lig 0,01. (Vand ved c. 20°). Endvidere er $\frac{L}{d^2}$ sat lig 20. Tabellen belyser altsaa alle de Tilfælde, der giver $\frac{L}{d^2}$ denne Værdi, f. Eks.

$d = 0{,}4 \qquad L = 3{,}0$
$d = 0{,}2 \qquad L = 0{,}8$
$d = 0{,}1 \qquad L = 0{,}2$

O. S. V.

$h - h_0$	$\frac{V_0^2 - V^2}{V^2}$	$\frac{d^2 V^2}{dh^2} \Big/ \frac{dV^2}{dh}$
1 cm	0,288	0,0720
2 -	0,204	0,0255
4 -	0,144	0,0045
10 -	0,091	0,0023

Tabellen viser navnlig, hvor overordentlig lidt V^2-h-Kurven maa ventes at ville krumme, hvis V_0^2-h-Kurven er retliniet. Indenfor det Interval, hvor Forsøg kan udføres — *ab* i Fig. 13 — vil Krumningen meget let kunne unddrage sig Opmærksomheden. Extrapolerer man nu retliniet til Skæringen med Abscisseaksen, ses det, at denne Extrapolation maa give en falsk — og for stor — Værdi for h_0.

Det er givet, at man maa tilskrive Gnidningen i Straalehullet Skylden for en Del af den tilsyneladende Kontraktion, d. v. s. den Del af den effektive Kontraktion, der ikke afspejler en virkelig Sammentrækning af Straalen. Vi har nu set, at Gnidningen tillige har den Virkning at nedsætte den virksomme Trykhøjde med et Beløb, som man ved mindre Hastigheder kan gøre fuldstændig Rede for gennem Poiseuilles Udtryk. I den reducerede V^2-h-Kurve er Indflydelsen paa Trykhøjden elimineret, men Gnidningens Bidrag til den effektive Kontraktion er paa langt nær ikke udskilt. Det fremgaar med stor Tydelighed af de nedenfor anførte Iagttagelser over den effektive Kontraktion for Vandstraaler ved en Række Boringer i Paraffin.

Forsøg	Boringens Diameter	Boringens Længde	Effektiv Diameterkontraktion	
			ukorrigeret	korrigeret
Nr. 57	1,648 mm	0,5 mm	6,7 %	6,5 %
- 58	1,567 -	2,0 -	4,5 -	3,7 -
- 59	1,677 -	1,0 -	6,4 -	6,2 -
- 60	1,475 -	4,0 -	10,4 -	8,6 -
- 61	1,535 -	8,0 -	15,4 -	12,6 -
- 62	1,409 -	8,0 -	15,0 -	14,8 -
- 63	1,993 -	8,0 -	9,7 -	8,2 -

Den effektive Kontraktions Variation med Længden gør sig, som det vil ses, omtrent lige saa stærkt gældende efter som før Korrektionen.

Det er ikke uden Interesse at lægge Mærke til, at medens man ved Bestemmelse af Vædskers indre Gnidning paa Grundlag af Poiseuilles Formel bruger Torricellis Udtryk til at korrigere Udstrømningstrykket for den Del, der bliver omsat i levende Kraft, benytter man ved den nu angivne Metode til Bestemmelse af Overfladespænding — hvilken Metode er grundet paa Torricellis Udstrømningslov — Poiseuilles Udtryk til at rette Trykhøjden for den Del, der medgaar til Overvindelse af den indre Gnidning.

Det foreliggende Arbejde er udført med Støtte fra Carlsbergfondet, hvis Bestyrelse jeg herved bringer min bedste Tak.

Færdig fra Trykkeriet d. 11. December 1919.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. II, 9.

ÜBER DEN NACHWEIS VON CHLORID NEBEN BROMID

VON

CARL FAURHOLT

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HOST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1919

Wie bekannt sind Silberchlorid, Silberbromid und Silberjodid in Wasser und verdünnten Säuren unlöslich; das Silberchlorid ist sowohl in Ammoniak als auch in Ammoniumsesquicarbonatlösung löslich; das Silberbromid ist in Ammoniak löslich, aber in einer Ammoniumsesquicarbonatlösung praktisch unlöslich; das letztere ist auch der Fall mit Silberjodid in Ammoniak, in einer Ammoniumsesquicarbonatlösung ist Silberjodid ebenfalls unlöslich.

Auf diese verschiedenen Grade der Löslichkeit fusst eine Methode, nach welcher man Chlorid, Bromid und Jodid nebeneinander nachweisen kann. Diese Methode gibt als erster HAGER an[1]; in einer etwas modificierten Gestalt wird sie in mehreren unserer hiesigen Unterrichtslaboratorien angewandt. Die Silberhaloide werden mit einem Überschuss von Silbernitrat ausgeschieden; durch wiederholtes Schütteln des Bodensatzes mit einer concentrierten Lösung von Ammoniumsesquicarbonat wird das Silberchlorid gelöst; mittels Ammoniak (ca. 4 n) wird auf dieselbe Weise das Silberbromid gelöst, während das Silberjodid ungelöst bleibt. Nach Ansäuerung der ammoniakalischen Lösungen mit Salpetersäure werden die Silberhaloide wieder ausgeschieden. Da die Silberhaloide mitein-

[1] Pharm. Centralhalle **12** (1871) pag. 42; Zeitschrift f. anal. Chemie **10** (1871) pag. 341.

ander aber feste Lösungen bilden[1] und es deshalb nicht so leicht ist, sie vollständig voneinander zu trennen, ist es notwendig, das Silberhaloid, welches an der Stelle des Silberbromides ausgeschieden wird, mit Zink und verdünnter Schwefelsäure zu behandeln; im Filtrat wird alsdann das Bromidion mittels Chlorwasser und Schwefelkohlenstoff nachgewiesen. In analoger Weise wird das Jodidion in dem in Ammoniumsesquicarbonat und Ammoniak ungelösten Reste nachgewiesen.

Das Silberchlorid jedoch ist in Ammoniumsesquicarbonat ziemlich schwerlöslich, und der Nachweis von Chlorid deshalb nicht ganz befriedigend. Da ohnedies eine Lösung von Ammoniumsesquicarbonat — je nach der Beschaffenheit des eingekauften Salzes — eine recht variierende Zusammensetzung besitzen kann, entstand in unserem Laboratorium der Wunsch ein anderes Lösungsmittel ausfindig zu machen, durch welches erstens eine effectivere Scheidung des Silberchlorids vom Bromid durchzuführen war, und welches zweitens leicht zu reproducieren war.

Warum löst sich denn Silberchlorid überhaupt in Ammoniumsesquicarbonatlösung, war die erste Frage, die einer Beantwortung harrte. Ammoniumsesquicarbonatlösung enthält sowohl saures Ammoniumcarbonat wie Ammoniumcarbaminat; in wässriger Lösung sind diese Salze in einem gewissen Grade in Ammoniak und in den freien Säuren hydrolysiert. Aller Wahrscheinlichkeit nach muss es diese geringe Menge von freiem Ammoniak sein, welche die Lösung bewirkt, und in diesem Falle müsste man ebensogut Ammoniakwasser von derselben Concentration wie die des freien Ammoniaks in der Ammoniumsesquicarbonatlösung anwenden können. Der Vorteil hierbei liegt an

[1] Zeitschrift f. anorg. Chemie **24** (1900) pag. 1.

der Hand: man ist im Stande, immer ein Reagens von constanter Beschaffenheit herzustellen — eine effectivere Scheidung als vermittels der Ammoniumsesquicarbonatlösung wird jedoch nicht erreicht; eine solche liesse sich aber erzielen, wenn man das Ammoniak mit Silbernitrat versetzte.

Wenn Silbernitrat in einer Menge zugesetzt wird, die von derselben Grössenordnung ist wie die der Löslichkeit des Silberchlorids, so wird hierdurch die Löslichkeit des Silberchlorids nur in geringem Grade beeinträchtigt, während die Löslichkeit des Silberbromids sehr verringert wird; man wird dadurch auch in den Stand gesetzt, bedeutend grössere Ammoniakconcentrationen anwenden zu können, und bekommt in dieser Weise einen Ersatz des Ammoniak-Puffergemisches, das eine Ammoniumsesquicarbonatlösung enthält. Die nachstehenden Versuche knüpfen an diesen Gedanken an.

Bodländer und Fittig[1] haben die Löslichkeit des Silberchlorids und Silberbromids in Ammoniak und ausserdem die Löslichkeit des Silberchlorids in Ammoniak mit ein wenig Silbernitrat versetzt untersucht, und auf diese und andere Versuche fussend, haben die genannten Forscher die Theorie der Lösung der Silberhaloide in nicht allzusehr concentriertem Ammoniak ermittelt.

Bei der Lösung des Silberchlorids in Ammoniak verlaufen die Processe, wie folgt:

$$AgCl \rightarrow Ag^{+} + Cl^{-}$$

und $$Ag^{+} + 2\,NH_3 \rightarrow Ag\,(NH_3)_2 \ .$$

Diesen Processen entsprechend ergeben sich die Massenwirkungsgleichungen:

[1] Zeitschrift f. physik. Chemie **39** (1902) pag. 597.

$$\frac{[Ag^+]\,[Cl^-]}{[AgCl]} = k'_1. \qquad (1)$$

und

$$\frac{[Ag(NH_3)_2^+]}{[Ag^+]\,[NH_3]^2} = k. \qquad (2)$$

Ist die Silberchloridlösung gesättigt, geht die Gleichung (1) in die folgende über:

$$[Ag^+]\,[Cl^-] = k_1. \qquad (3)$$

Durch Combination der Gleichungen (2) und (3) erhält man

$$\frac{[Ag(NH_3)_2^+]\,[Cl^-]}{[NH_3]^2} = k \cdot k_1 = K_1.$$

Für das Silberbromid hat eine ganz analoge Gleichung Gültigkeit. Durch Einführen der Constantenwerte in diese Gleichungen liess die Löslichkeit des Silberchlorids und Silberbromids sich annähernd berechnen, und die Resultate zeigten auch, wie nach den Grössen der beiden Löslichkeitsprodukte zu erwarten war, dass wenig Silbernitrat nur einen geringen Einfluss auf die Löslichkeit des Chlorides ausübte, einen bedeutenden dagegen auf die des Bromides. Diese Berechnungen als Ausgang benutzend, wurden etliche Lösungsmittel ausgewählt, deren Lösungswirkungen untersucht wurden.

1. Bei einigen vorläufigen Versuchen zeigte sich indes, dass das Silberchlorid und namentlich das Silberbromid nach kräftigem 3 Minuten langem Schütteln — also in derselben Weise wie die Operation in der qualitativen Analyse ausgeführt wird — mit den verdünnteren Lösungsmitteln zum Teil durch das Filter gerissen wurde. Dieses wird durch Zusatz von ein wenig Elektrolyt verhindert. Kaliumnitrat wurde als solches gewählt, und sämtliche

Lösungsmittel sind folglich mit etwas (0,25 $^{Mol}/_{l}$) Kaliumnitrat versetzt.

Das Kaliumnitrat ist durch Umkristallisation gereinigt, bis es chlorfrei war. Das Ammoniak war von den Pyridinbasen befreit.

Das Silberchlorid wurde unmittelbar vor dem Gebrauche hergestellt; es geschah dies durch Fällen von Salzsäure mit Silbernitrat, Auswaschen des Bodensatzes und schliesslich Entfernen des meisten Wassers mittels der Saugpumpe. Die Pasta, die ca. 25% Wasser enthielt, verteilte sich leicht in den Lösungsmitteln.

Auch das Silberbromid wurde unmittelbar vor dem Gebrauche hergestellt; es geschah durch Fällen einer ammoniakalischen Kaliumbromidlösung mit Unterschuss von Silbernitrat. Der Bodensatz wurde nach einem Auswaschen mit schwachem Ammoniak 3 Stunden rotiert, wonach er wieder ausgewaschen wurde. Zuletzt wurde das meiste Wasser abgesogen. Das Silberbromid wurde also ca. 10 Stunden alt, ehe es mit den Lösungsmitteln in Berührung kam.

Die endgültigen Löslichkeitsversuche wurden bei 18° durchgeführt; es wurde ein Überschuss des Silberhaloids 12 Stunden mit dem Lösungsmittel rotiert. Einzelne Versuche, in welchen die Rotation 24 Stunden dauerte, zeigten, dass 12 Stunden hinreichend waren. Besondere Vorsichtsmassregeln gegen die Einwirkung des Lichts wurden nicht getroffen, intensives Tageslicht wurde jedoch vermieden. Durch Ansäuren von in der Regel nur 100 cm^3 Filtrat mittels Salpetersäure wurde das Silberhaloid ausgefällt, wonach es entweder als solches oder, wenn das Silberhaloid weniger als 0,1 g wog, nach erfolgter Reduction zu metallischem Silber gewogen wurde. Die kleinste

Silbermenge, welche gewogen wurde, war 0,01 g. In einigen der Lösungsmittel wurde so wenig Silberbromid gelöst, dass eine gewichtsanalytische Bestimmung unmöglich war, es sei denn, dass man mehrere Liter des Lösungsmittels anwendete; die gelösten Mengen von Silberbromid wurden in diesem Falle bestimmt, oder vielmehr es wurde versucht, sie durch Nephelometrie zu bestimmen. Des Vergleiches halber wurden auch in den concentrierteren Silberbromidlösungen — nach passender Verdünnung — nephelometrische Bestimmungen vorgenommen.

Die Versuchsresultate sind in der Tabelle I aufgeführt:

TABELLE I

Nr.	Zusammensetzung des Lösungsmittels			Silberchlorid gewichtsanalytisch		Silberbromid gewichtsanalytisch		Silberbromid nephelometrisch	
	NH_3 Mol/l	$AgNO_3$ Mol/l	KNO_3 Mol/l	g/l	Mol/l · 10^4	g/l	Mol/l · 10^4	g/l	Mol/l · 10^4
1	0,25	0	0,25	2,373	165,6	0,181	9,64	0,165	8,80
2	0,50	0	0,25	4,941	344.8	0,353	18,82	0,338	18,00
3	1,00	0	0,25	10,350	722,1	0,714	38,01	0,789	42,00
4	0,25	0,01	0,25	1,667	116.3			0,008*	0,40*
5	0,50	0,01	0,25	4,209	293,1	0,061	3,24	0,041	2,20
6	1,00	0,01	0,25	9,624	671,6	0,251	13,35	0,244	13,00
7	0,50	0,05	0,25	2,061	143,8			0,005*	0,25*
8	Ammoniumsesquicarbonat 80g/l			0,218	15,2			0,009*	0,50*

* Diese Angaben sind zu niedrig, wie sich aus dem folgenden ergeben wird.

Die Tabelle I zeigt, dass die zugesetzten Silbernitratquanta nur einen geringen Einfluss auf die Löslichkeit des Silberchlorids haben, einen grossen dagegen auf diejenige des Silberbromids. In der Tabelle II findet sich die Gleichgewichtsconstante K_1 und die analoge Constante für Silberbromid K_2, unter Voraussetzung der vollständigen

Ionisation berechnet. Dass ein wenig des Ammoniaks als Ammoniumhydroxyd zugegen ist, ist auch nicht berücksichtigt; aber von der totalen Ammoniakconcentration sind doppelt so viele Moleküle, als Silberatome in der Lösung sind, subtrahiert. Die in dieser Weise reducierte Ammoniakmenge ist, da das Ammoniak nicht ganz Henrys Gesetz folgt, — wie von Bodländer und Fittig angegeben — mit $(1+0{,}08\,n)$ multipliciert; n gibt hier die Normalität an.

TABELLE II

Nr.	Zusammensetzung des Lösungsmittels			Silberchlorid $K_1 \cdot 10^5$	Silberbromid $K_2 \cdot 10^5$	
	NH_3 Mol/l	$AgNO_3$ Mol/l	KNO_3 Mol/l	gewichtsanalytisch	gewichtsanalytisch	nephelometrisch
1	0,25	0	0.25	563	1,45	1,22
2	0,50	0	0,25	598	1,33	1,21
3	1,00	0	0,25	624	1,26	1,54
4	0,25	0,01	0,25	570		0,73
5	0,50	0,01	0,25	607	1,35	0,90
6	1,00	0,01	0,25	635	1,38	1,32
7	0,50	0,05	0,25	634		0,73
					Mittelwert	
					1,35	

Die Übereinstimmung zwischen den Werten von K_1 ist als gut zu betrachten, wenn man im Auge behält, dass mit der vollständigen Ionisation gerechnet wird. Die Übereinstimmung zwischen den Werten von K_2 ist ebenfalls gut, die drei Werte ausgenommen, welche von den drei verdünntesten Silberbromidlösungen herrühren und auf Grund nephelometrischer Bestimmungen berechnet sind.

Eine von den verdünnten Lösungen (Versuch Nr. 5), von welcher auch eine gewichtsanalytische Bestimmung in 500 cm^3 Filtrat vorliegt, zeigt, dass die nephelometrisch

bestimmten Werte in den verdünnten Lösungen zu klein sind; diese Bestimmungen sind jedoch mitgenommen, da sie in ausgezeichneter Weise die Grössenordnung der gelösten Silberbromidmengen illustrieren. Dass drei der nephelometrischen Bestimmungen zu niedrig ausgefallen sind, beruht sicher darauf, dass die Lösung von dem ungelösten Silberbromid filtriert worden ist, und dass das Filtrierpapier etwas von dem gelösten Bromide adsorbiert hat. Es liess sich in der Tat nachweisen, dass das Filtrierpapier eine nicht geringe Menge Bromid adsorbiert. Von einer ammoniakalischen, sehr verdünnten Silberbromidlösung wurden 75 cm^3 durch Papier filtriert und das Filtrat in drei gleich grosse Teile fractioniert. Nach Ansäuerung mit Salpetersäure zeigte die erste Fraction eine Opalisierung, die nur ca. halb so gross war wie diejenige, welche in 25 cm^3 unfiltrierter Lösung entstand; die Opalescenz in der dritten Fraction war grösser als in der ersten, aber doch nicht so gross wie in der unfiltrierten Lösung. Bei den concentrierteren Lösungen ist der adsorbierte Bruchteil so gering, dass die Löslichkeitsbestimmungen dadurch nicht beeinträchtigt werden. Mit dem gefundenen Mittelwerte von K_2, $1{,}35 \cdot 10^{+5}$, lässt sich die richtige Löslichkeit des Silberbromids in den Lösungsmitteln Nr. 4 und 7 berechnen. Die Werte finden sich in der Tabelle III.

TABELLE III

Nr.	Zusammensetzung des Lösungsmittels			Silberbromid berechnet	
	NH_3 Mol/l	$AgNO_3$ Mol/l	KNO_3 Mol/l	g/l	Mol/l · 10^4
4	0,25	0,01	0,25	0,014	0,74
7	0,50	0,05	0,25	0,009	0,50

In der Tabelle IV ist angegeben, wie vielmal Silberchlorid in den verschiedenen Lösungsmitteln leichter löslich ist als Silberbromid. Die Zahlenwerte dieses Verhältnisses sind nach den gewichtsanalytischen Bestimmungen berechnet, wo solche nicht existieren, nach den berechneten Löslichkeiten in der Tabelle III.

TABELLE IV

Nr.	Zusammensetzung des Lösungsmittels			Mol Ag Cl. / Mol Ag Br.
	NH_3 Mol/l	$AgNO_3$ Mol/l	KNO_3 Mol/l	
1	0,25	0	0,25	17,2
2	0,50	0	0,25	18,3
3	1,00	0	0,25	19,0
4	0,25	0,01	0,25	157
5	0,50	0,01	0,25	90
6	1,00	0,01	0,25	50
7	0,50	0,05	0,25	288

Die Tabelle IV zeigt, dass das Verhältnis zwischen den molaren Löslichkeiten von Silberchlorid und Silberbromid in reiner, schwacher Ammoniaklösung ca. 18 ist. Dieses Verhältnis ist demjenigen reiner, wässriger Lösungen gleich, und sollte es infolge der Theorie auch sein, wenn die Bodensätze, mit welchen die ammoniakalischen Lösungen in Gleichgewicht sind, aus reinem Silberchlorid oder Silberbromid bestehen. Wenn die Ammoniaklösung Silbernitrat enthält in Mengen, welche von derselben Grössenordnung sind wie die Löslichkeit des Silberchlorids, steigt das Verhältnis zwischen der Löslichkeit des Silberchlorids und Silberbromids bedeutend über 18.

2. Vergleichshalber sind mit einer Ammoniumsesquicarbonatlösung einzelne Versuche angestellt — Nr. 8 in der Tabelle I. Die Lösung, welche bei der Anwendung einige Tage alt war, enthielt 80 g Salz pro Liter. Der

nephelometrisch bestimmte Löslichkeitswert des Silberbromids ist zu niedrig. Von der Annahme ausgehend, dass das Verhältnis zwischen der Löslichkeit des Silberchlorids und Silberbromids auch bei sehr kleinen Ammoniakconcentrationen ca. 17 ist, wird die Löslichkeit des Silberbromids in der obgenannten Ammoniumsesquicarbonatlösung zu 0,017 $^{g}/_{l}$ und die molare Concentration zu $0{,}89 \cdot 10^{-4}$ berechnet. Dieser berechnete Wert stimmt einigermassen mit einem experimentellen von Baubigny[1] überein.

3. Von den untersuchten Lösungsmitteln sind Nr. 4 und 7 diejenigen, welche die beste Trennung der Silbersalze zulassen. Welches von ihnen man anwenden will, ist einigermassen Geschmacksache. Das Lösungsmittel Nr. 7 verursacht eine etwas bessere Trennung als Nr. 4, und bei Nr. 7 spielt eine weniger vollständige Auswaschung der freien Säure eine weniger bedeutende Rolle als bei Nr. 4 infolge des grösseren Ammoniakgehaltes. Andererseits ist Nr. 4 ein billigeres Reagens.

Da Silberchlorid und Silberbromid, wie oben angedeutet, feste Lösungen miteinander bilden, darf man auf Grund der gefundenen Löslichkeitswerte des reinen Silberchlorids und Silberbromids nur schliessen, dass die Trennung mit einer silbernitrathaltigen Ammoniaklösung von zweckmässiger Zusammensetzung ca. 10mal so gut wie eine Trennung mit einer Ammoniumsesquicarbonatlösung sein wird, aber es ist nicht möglich, einen Schluss auf die Feinheit der Trennungsmethode zu machen. Um diese ausfindig zu machen, wurden etliche qualitative und grobe quantitative Löslichkeitsbestimmungen mit dem Lösungsmittel Nr. 4 (0,25 n NH_3, 0,01 n $AgNO_3$, 0,25 n KNO_3) vorgenommen. Diese Lösung wird nun in dem hiesigen Unterrichts-

[1] Comptes rendus **146** (1908) pag. 335.

laboratorium benutzt und wird einfach »Silbernitrat-Ammoniak« genannt.

0,2 g Kaliumbromid und wechselnde Mengen Kaliumchlorid wurden in Wasser gelöst und mit Überschuss von Silbernitrat gefällt. Nach erfolgter Auswaschung wurde der Bodensatz drei Minuten mit ca. 10 cm^3 Silbernitrat-Ammoniak geschüttelt. Ein Gehalt von 3—1 % Kaliumchlorid gab noch eine für einen angehenden Analytiker sichere und deutliche Chloridreaktion, der geübtere weist mit Leichtigkeit einen noch bedeutend geringeren Chloridgehalt nach.

Lösungen mit wechselnden Mengen von Kaliumbromid und Kaliumchlorid wurden mit Überschuss von Silbernitrat gefällt. Der Bodensatz wurde nach erfolgter Auswaschung 12 Stunden bei 18° mit 100 cm^3 Silbernitrat-Ammoniak, das $11{,}6 \cdot 10^{\div 4}$ Mol Silberchlorid lösen kann, rotiert. In dem Filtrate wurden die gelösten Silberchloridmengen gewichtsanalytisch bestimmt. Die Resultate sind in der Tabelle V zu finden.

TABELLE V

Die Lösung enthält:		Mol AgCl im Bodensatz vor der Behandlung mit Silbernitrat-Ammoniak	Mol AgCl in 100 cm^3 Silbernitrat-Ammoniak gelöst	% AgCl gelöst
KBr g	KCl g			
2,0	0,038	$5 \cdot 10^{\div 4}$	$1 \cdot 10^{\div 4}$	20
2,0	0,075	$10 \cdot 10^{\div 4}$	$1{,}5 \cdot 10^{\div 4}$	30
0,2	0,075	$10 \cdot 10^{\div 4}$	$5 \cdot 10^{\div 4}$	50

4. Aus der in Ammoniumsesquicarbonat gelösten Menge von Silberchlorid in Verbindung mit den in der Tabelle II berechneten Werten von K_1 ergibt sich, dass die molare Concentration des freien Ammoniaks in der Ammoniumsesquicarbonatlösung ca. 0,02 sein muss, unter der Voraus-

setzung, dass sich nicht andere Einflüsse bei der Lösung der Silberhaloide geltend machen. Dass die Ammoniakconcentration in der Tat von der genannten Grössenordnung ist, liess sich in zweierlei Weisen abschätzen.

Erstens ergab eine colorimetrische Wasserstoffionenbestimmung, dass $p_{H^+} = 8{,}1$ ist; folglich ist $p_{OH^-} = 14{,}2 - 8{,}1 = 6{,}1$. Bei 18° wird $[NH_3] = \frac{[NH_4^+][OH^-]}{1{,}75 \cdot 10^{-5}}$ sein.

Nach Einsetzen des Wertes der Hydroxylionconcentration ergibt sich $[NH_3] = [NH_4^+] \cdot 0{,}04539$.

Die Ammoniumionconcentration in der angewandten Ammoniumsesquicarbonatlösung, welche ca. $^1/_2$ molar ist sowohl rücksichtlich des sauren Ammoniumcarbonates wie auch des Ammoniumcarbaminates, ist jedenfalls etwas weniger als 1molar, sie lässt sich passend dem Ionisationsgrad des Natriumacetates in einer 1molaren Lösung gleichsetzen. Es ergibt sich hieraus:

$$[NH_3] = 0{,}525 \cdot 0{,}04539 = 0{,}024.$$

Die zweite Methode, nach welcher versucht wurde, die Grössenordnung der Ammoniakconcentration zu bestimmen, bestand darin, die Concentration desjenigen Ammoniakwassers zu bestimmen, welches dieselbe Ammoniaktension wie die Ammoniumsesquicarbonatlösung hatte. Diese wurde dadurch bestimmt, dass in gleich grossen Zeiträumen gleich grosse Volumina Luft teils durch die Ammoniumsesquicarbonatlösung, teils durch Ammoniaklösung in zweckmässiger Verdünnung gesogen wurden; das entweichende Ammoniak wurde in Säure aufgefangen und titriert. Bei jedem Versuche wurden 2mal 30 Liter Luft 2mal 25 Minuten durch 200 cm^3 Flüssigkeit gesogen. Die Resultate finden sich in der nachstehenden Tabelle VI.

TABELLE VI

	Ammoniumsesquicarbonatlösung	0,02n Ammoniak
cm^3 0,1-n NH_3, bei der 1sten Durchlüftung entwichen	3,2	1,9
» » » » » 2ten » »	5,4	1 8

Aus der Ammoniumsesquicarbonatlösung, welche schon beim Schütteln etwas Luftentwickelung gibt, entweicht auch Carbondioxyd und wahrscheinlich verhältnismässig mehr von diesem Gase als von dem Ammoniak. Hierdurch wird die Lösung ammoniakalischer, als sie in der Wirklichkeit ist, und dieses Verhältnis macht sich möglicherweise schon im Werte der ersten Durchlüftung geltend, sicher aber in demjenigen der zweiten. Die Zahlen in der Tabelle VI gestatten mit grosser Wahrscheinlichkeit den Schluss, dass die Ammoniumsesquicarbonatlösung höchstens 0,03 normal in bezug auf das Ammoniak sein kann, wahrscheinlich aber ist der Wert etwas geringer. Die durchlüfteten 200 cm^3 0,02-n Ammoniakwasser entsprechen 40 cm^3 0,1-n Ammoniakwasser, es ist also nur ein geringer Bruchteil des Ammoniaks, welcher entwichen ist.

5. Mit Silberchlorid und Silberbromid, welche entweder 8 Tage in feuchtem Zustande gestanden hatten oder ein paar Stunden bei 130° getrocknet worden waren, sind auch einzelne Versuche angestellt worden. Sie hatten dieselbe Löslichkeit wie die frischgefällten Salze.

6. Die Scheidung des Silberbromids und Silberjodids mittels Ammoniakwasser ist immer in der besten Weise verlaufen. Da die wenigen Bestimmungen, welche über die Löslichkeit des Silberjodids in Ammoniakwasser vorliegen, sehr divergieren, ist hier, der Vollständigkeit halber, eine

einzelne gewichtsanalytische Löslichkeitsbestimmung des Silberjodids vorgenommen. 1 Liter 4n Ammoniak, welches zugleich 0,25n in bezug auf Kaliumnitrat war, löste 0,05 g oder $2{,}15 \cdot 10^{-4}$ Mol Silberjodid. Das Silberjodid, das in ähnlicher Weise wie das Silberbromid dargestellt war, wurde weiss beim Schütteln mit Ammoniakwasser. Eine Analyse der Flüssigkeit und des feuchten Bodensatzes zeigte an, dass der weisse Bodensatz der Formel $5\,AgJ,\,2\,NH_3$ entsprach. Ist die Normalität des Ammoniakwassers geringer als ca. 3, bleibt das Silberjodid gelb.

Resultat: Es wird zum Nachweis von Chlorid neben Bromid nach Hagers modificierter Methode vorgeschlagen, anstatt einer Ammoniumsesquicarbonatlösung eine silbernitrathaltige Ammoniaklösung, welche zugleich Kaliumnitrat enthält, anzuwenden.

Die Lösung kann passend pro Liter enthalten:

0,25 Mol Ammoniak, 0,01 Mol Silbernitrat und
0,25 Mol Kaliumnitrat,

oder 0,50 Mol Ammoniak, 0,05 Mol Silbernitrat und
0,25 Mol Kaliumnitrat.

Die vorliegende Arbeit ist in dem chemischen Laboratorium der Copenhagener »Kongelige Veterinær- og Landbohøjskole« ausgeführt. Es ist mir eine angenehme Pflicht, dem Leiter dieses Laboratoriums, meinem hochverehrten Chef, Hrrn. Professor, Dr. phil. N. Bjerrum, meinen besten Dank auszusprechen für die mannigfache Anregung und die stete Hülfe, welche er mir während dieser Arbeit zuteil werden liess.

Færdig fra Trykkeriet den 8. December 1919.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. II, 10.

ON THE SOLUBILITY OF SALTS IN SALT SOLUTIONS

STUDIES ON SOLUBILITY I

BY

J. N. BRØNSTED

WITH 10 FIGURES IN THE TEXT

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HOST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1919

1. Introduction.

The fundamental principles for the treatment of the solubility equilibria in salt solutions are pointed out in the well known works of VAN'T HOFF and REICHER[1], W. NERNST[2] and ARTHUR A. NOYES[3]. Through these works it was shown that the important conception of the solubility product was able to represent the solubility phenomena in concordance with the VAN'T HOFF theory of the dilute solution and the ARRHENIUS theory of the electrolytic dissociation of salts. Indications, however, are found already in these first papers, of the fact, that the assumption of the constancy of the solubility product, made in the quoted papers and afterwards admitted as a general rule, is able to account only approximately for the experimental results.

Further investigations by ARRHENIUS[4] and by STIEGLITZ[5] showed the concentration of the unionized portion of the saturating salt to be not even approximately constant as required by the theory of solutions, this concentration decreasing rapidly by increasing concentration of another salt with an ion in common with the saturating salt.

[1] Zeitschrift f. physikalische Chemie 3, 482 (1889).
[2] Ibidem 4, 372 (1889).
[3] Ibidem 6, 241 (1890).
[4] Zeitschr. f. physik. Chemie 31, 224 (1899).
[5] Journ. Am. Chem. Soc. 30, 946 (1908).

A considerable amount of work concerning the solubility of salts has been carried out after this, of which notably the important investigations of Noyes and his coworkers[1] deserve special mention. Through this elaborate work experimental results have been brought about to corroborate the view of Arrhenius named above, and showing, furthermore, the solubility product in mixed salt solutions to rise considerably with increasing concentration.

To obtain this result the authors apply the method of the electric conductance of the solutions, assuming the conductance ratio to be an accurate measure of ionization in pure salt solutions. By introducing the Arrhenius theory of the isohydric solution, and the empirical law which has been found to express the change of the ionization of single salts, an expression is reached which is assumed to govern the equilibria occurring in the mixed saturated solutions.

The results obtained in this way are dependent on the precision of the fundamental assumption, of the conductance ratio being an accurate measure for the degree of dissociation. There has been much discussion respecting the value of this assumption, and we are justified at least in considering it as highly doubtful[2], since the underlying presupposition that the mobility of the ions is constant is shown to be incorrect. Also for other reasons we cannot adopt the conductance method as fitted to account for the constitutions of salt solutions, as it will appear from an inspection of the solubility data. It may be stated f. inst. that even when a fall in the concentration of the

[1] Journ. Am. Chem. Soc. 33, 1643 (1911). 33, 1807 (1911).

[2] See esp. G. N. Lewis, Journ. Am. Chem. Soc. 34, 1631, (1912) and P. Hertz, Ann. d. Physik. [4] 37, 1 (1912).

unionized portion by addition of another salt may be expected, from the phenomenon well known as the "salting out" effect, this fall as calculated in the cited papers proves to be of quite a different order of magnitude than in the case of a non-electrolyte. Thus a 0.05 kalium nitrate solution, possessing a salting out effect of less than 1 per cent towards non-electrolytes, exhibits an effect more than 30 times as great in the case of undissociated thalleous chloride, computed on the base of the conductance method. Furthermore, a satisfactory explanation of the increasing solubility found in many cases in salts on addition of other homoionic salts can hardly be afforded from this point of view. By NOYES and BRAY[1] attention is called to the peculiar behaviour of a moderately soluble uni-bivalent salt on addition of a salt with a common bivalent ion. Neither is this phenomenon easily understood from the standpoint of the ordinary theory. We are justified therefore, in consideration of all of these discrepancies, in concluding that the conductance method affords no serviceable base for a theory of solubility. It seems desirable therefore to attempt an interpretation of these phenomena from quite a different point of view.

The attempt made in the present paper at a new treatment of the solubility problem makes no pretence of explaining all peculiar characteristics of the various solubility curves. It is intended only to show, that the multiplicity of types of curves found by employment of hetero- as well as homoionic solvents can be accounted for in a simple way and on thermodynamic base by introducing an assumption at any rate approximately true for mixed salt solutions.

[1] Journ. Am. Chem. Soc. 33, 1643 (1908).

Owing to the preliminary nature of this paper an account of a considerable experimental material, which has been produced in this laboratory to test the general scope of the new points of view, will be postponed for later publications. Only a small number of experiments will be found in the following sections of this article to elucidate some important conclusions. A closer theoretical treatment of the problem will likewise appear in a subsequent part of this work.

2. Thermodynamic Functions Applicable to Solutions.

The thermodynamic properties of a dissolved substance can be expressed by means of the chemical potential π, this quantity in the case of the validity of the simple gas laws being given by

$$\pi = RT \ln c + i_c \qquad (1)$$

for one mol of the solute, where c is the concentration or number of mols of the solute per liter, and i_c a constant for the same solvent.

If the gas laws do not hold for the solution, this equation may be displaced by

$$\pi = RT \ln \xi_c + i_c \qquad (2)$$

where ξ_c denotes the activity of the solute, a conception introduced by G. N. Lewis[1] and conveniently defined by this very equation.

The absolute value of the activity is not determined by (2) and has commonly no significance. It may be established by definition by putting $\xi_c = c$ of the substance in the pure gaseous state at indefinite dilution. For our purpose, however, it may be more suitable to put

[1] Proc. Am. Acad. 43, 259 (1907).

$\xi_c = c$ for the substance in the employed solvent by indefinite dilution. As a more complete definition of the activity we then have

$$\pi - \pi_\infty = RT \ln \frac{\xi_c}{c_\infty} \qquad (3)$$

where π_∞ and c_∞ indicate the chemical potential and the concentration of the substance by indefinite dilution in the solvent concerned.

The ratio of the activity to the concentration:

$$\frac{\xi_c}{c} = f \qquad (4)$$

is called the activity coefficient[1]. The value of this is constant when equation (1) holds good, but otherwise varies with varying concentration. When ξ_c is given by (3) the said constant value will be unity.

While the alterations in ξ_c by changing concentration as seen from equation (2) and (3) is a merely experimental quantity, no hypothesis whatever being involved for its theoretical determination, the values of the alterations in c, however, will usually depend upon the method employed for the computation of this quantity. This uncertainty of course also affects the activity coefficient. Different values must be attributed to f, depending on the method employed for the determination of c.

This uncertainty will be removed by introducing for c the stoichiometrically determined concentration of the dissolved substance instead of the "true" concentration. The activity coefficient thereby becomes a thermodynamic quantity like the activity itself. Defined in this way it may be suitably termed the "stoichiometric activity coefficient".

[1] Noyes a. Bray. Journ. Am. Chem. Soc. 33, 1646 (1911).

From the nature of equation (1) it appears that we need not necessarily employ the c-scale to express the concentration in this equation. By substituting x or the mol fraction for c we obtain

$$\pi = RT \ln x + i_x \tag{5}$$

analogous to (1) and

$$\pi = RT \ln \xi_x^{\nu} + i_x \tag{6}$$

analogous to (2). Instead of (3) and (4) me way write

$$\pi - \pi_\infty = RT \ln \frac{\xi_x}{x_\infty} \tag{7}$$

and

$$\frac{\xi_x}{x} = f. \tag{8}$$

The activities ξ_c and ξ_x are capable of representing the thermodynamic properties of the solution equally well. In dividing by the corresponding concentrations c and x, the same activity coefficient f will result. As in the case of applying the c-scale we have also here to distinguish between the true and the stoichiometric activity coefficient.

An equation of the same significance for dilute solutions as (1) and (5) is arrived at by expressing the chemical potential of the solvent in dependence of the concentration. If by π_0 and π_{00} are indicated the chemical potentials of one mol of the solvent at the concentration x and indefinite dilution respectively, we have for dilute solutions if the gas laws hold good

$$\pi_0 - \pi_{00} = -RTx. \tag{9}$$

The magnitude of π and x depends on the molal weights M_0 and M, attributed to the solvent and the solute respectively. As to M_0 this may, strictly speaking, be

chosen arbitrarily but is most conveniently taken as equal to the formula weight of the substance. By fixing M_0 we shall find that the value of M is also determined according to equation (9). The gas laws being valid for the solution means precisely that M, derived in this way from (9), keeps a constant value when the concentration changes.

If the gas laws are not valid, (9) must be replaced by

$$\pi_0 - \pi_{00} = -RT\eta_x \tag{10}$$

holding for the solution irrespective of its concentration. η_x may be termed the osmotic concentration and is, like ξ, a merely thermodynamic quantity. Putting

$$\frac{\eta_x}{x} = \varphi \tag{11}$$

this ratio called the osmotic coefficient[1] proves to be a thermodynamic quantity also when x is the stoichiometric concentration calculated from the above values of M and M_0. Accordingly we have in this case also to distinguish between the true and the stoichiometric coefficient.

Expressing the concentration in the c-scale, (10) and (11) are replaced by the corresponding formulae

$$(\pi_0 - \pi_{00})\frac{c}{x} = -RTc \tag{12}$$

and

$$(\pi_0 - \pi_{00})\frac{c}{x} = -RT\eta_c. \tag{13}$$

Furthermore we obtain

$$\frac{\eta_c}{c} = \varphi. \tag{14}$$

[1] Bjerrum. Zeitschrift f. Elektrochemie 24, 321 (1918).

By the equations above a series of functions: ξ_c, ξ_x, η_c, η_x, f and φ are introduced for the purpose of thermodynamic treatment of solutions. Of course, the introduction of these quantities affords no actual progress in the treatment in comparison with the application of the chemical potential or the affinity. All these quantities are correlated thermodynamically and are thermodynamically of the same worth, and only for the sake of convenience or brevity one of them is preferable to another. It will be found, however, as especially pointed out by Bjerrum[1], that in the case of dilute salt solutions, which we are going to consider in the following, the activity coefficient and the osmotic coefficient will prove to be of peculiar value for the theoretical treatment.

The correlations between the functions introduced may be deduced by means of the fundamental equation of W. Gibbs:

$$x d\pi + d\pi_0 = 0 \qquad (15)$$

which holds for dilute solutions, the same equation that in the form:

$$x \frac{dA_1}{dx} + (1 - x) \frac{dA_2}{dx} = 0$$

has proved of value in affinity determinations. Introducing here (2), (6), (10), and (13) we obtain the equation

$$x\, dln\, \xi_x = x\, dln\, \xi_c = d\eta_x = \frac{x}{c} d\eta_c \qquad (16)$$

which expresses the relation betveen the activities and the osmotic concentrations.

If more than one substance is present in dilute solution in the solvent, we have to apply the more general formula

[1] l. c.

$$x_1 d\pi_1 + x_2 d\pi_2 + \ldots + d\pi_0 = 0 \qquad (17)$$

from which we obtain

$$\left.\begin{aligned} \Sigma x_1 \, dln\, \xi_{x_1} &= d\eta_x \\ \Sigma c_1 \, dln\, \xi_{c_1} &= d\eta_c \end{aligned}\right\} \qquad (18)$$

and

By insertion of (4), (8), (11), and (14), equation (18) gives us

$$\Sigma x_1 \frac{dln f_1}{dx} = x \frac{d\varphi}{dx} - (1 - \varphi) \qquad (19)$$

and

$$\Sigma c_1 \frac{dln f_1}{dc} = c \frac{d\varphi}{dc} - (1 - \varphi) \qquad (20)$$

where

$$\Sigma x_1 = x \quad \text{and} \quad \Sigma c_1 = c.$$

Putting $x_2 = x_3 = \ldots = 0$, equation (20) is transformed to the equation given by BJERRUM[1] and valid for binary mixtures.

As we are here going to deal with solutions containing two or more salts for the purpose of determining the solubility equilibrium, equations (18)—(20) may be of special importance.

3. The Theory of Solubility of Salts in Salt Solutions.

In a saturated solution of a binary salt the chemical potential of the saturating salt must possess a constant value equal to the sum of the chemical potentials of its ions. Denoting by π, π' and π'' the potential of the salt, the cation and the anion respectively, we have

$$\pi = \pi' + \pi'' = \text{constant}, \qquad (21)$$

or according to (2)

$$\xi' \xi'' = k \qquad (22)$$

[1] l. c.

which shows the activity product to be constant in a saturated solution. Inserting the activity coefficient by means of (4) we obtain

$$c' c'' f' f'' = k. \tag{23}$$

As in this equation c indicates the molal concentration, f must be the stoichiometric activity coefficient. The alteration of f in the case of addition of another salt must cause the product $c' c''$ — the stoichiometric solubility product — to change too with the concentration of the added salt, and this change will depend solely upon the alteration in f. The laws according to which the alteration in f takes place must therefore also govern the changing solubility product by addition of extraneous salts.

We shall now for the further treatment of the problem introduce the hypothesis referred to in the introduction. We shall assume that the activity coefficient in a mixed salt solution is the same for any ion of the same type. As will appear from the following this assumption will markedly simplify the problem before us. Then, since we are only dealing with comparatively sparingly soluble salts, the saturated solution will be either so dilute as to annul, at least approximately, the individual character of any salt, or it will consist nearly completely of the solvent alone. It will be possible therefore to calculate the activity coefficient by considering only the properties of solutions of single salts.

It is necessary to emphasize, however, the approximative and provisional nature of the above hypothesis. It is introduced here to show only that the general features of the solubility curves may be accounted for on a thermodynamic base by its means without further assumptions

as to the constitution of the solutions, and especially regardless of the degree of dissociation, contingently occurring in the solutions, and hitherto introduced for the treatment of saturated solutions by means of the conductance ratio.

For the calculation of the activity coefficients we can use the results of NOYES and FALK[1], who have shown, that the freezing point data for a number of binary salts can be expressed by the equation

$$2 - i = K \sqrt[3]{c}, \tag{24}$$

where i is the VAN'T HOFF factor, c the equivalent concentration, and K a constant characteristic of each salt. Introducing $2\alpha = K$, (24) may be written

$$\varphi = 1 - \alpha \sqrt[3]{c} \tag{25}$$

from which we obtain by means of (20) and (17)

$$\log^{10} f = -1{,}737\, \alpha \sqrt[3]{c} = -a \sqrt[3]{c}. \tag{26}$$

If this value be inserted in (23), we have according to our assumption

$$\log (c'\, c'') = 2a \sqrt[3]{c_t} + \text{const.} \tag{27}$$

c_t denoting the total salt concentration in the saturated solution given by $c_t = c + s$.

Properly speaking, on the above assumption this relation (27) will hold good only if the solubility is slight when compared with the concentration of the solvent. Since the values of K, as shown by NOYES and FALK, vary only moderately within the range of salts of the same type, and since the presupposition made by assuming the f values to be equal is only of approximate nature, no essential

[1] Journ. Am. Chem. Soc. 32, 1011 (1910).

restriction in the applicability of (27) will result if the requirement above as to the solubility is dispensed with.

In the case of uni-univalent salts the coefficient a as computed from the data of Noyes and Falk proves to oscillate around a value not far from $^1/_3$. This figure may therefore be accepted as an average value and introduced in (27) to express the solubility in this case. If the valencies of the ions are higher than unity, a will increase considerably.

From the freezing point data a value for a may be derived which for bi-bivalent salts is approximately four times the a value for uni-univalent salts. This is in concordance with theoretical considerations which render it likely that the a values increase proportionately to the square of the valency. In the case of bi-bivalent and tri-trivalent salts the values $a = {}^4/_3$ and $a = 3$ respectively may be adopted to express the solubility of such salts when introduced in (27).

For the application of this formula we have to distinguish between the cases of the solvent having one ion or none in common with the saturating salt. We may firstly consider the case of the employment of a heteroionic solvent.

4. Heteroionic Solvents.

In this case c' and c'' in equation (27) are equal to the solubility or

$$\log s = a \sqrt[3]{c_t} + \text{const.} \tag{28}$$

Introducing here

$$\log s_0 = a \sqrt[3]{s_0} + \text{const.} \tag{29}$$

we obtain

$$\log \frac{s}{s_0} = a \left(\sqrt[3]{c_t} - \sqrt[3]{s_0} \right). \tag{30}$$

By means of this relation the solubility ratio s/s_0 can be calculated for varying values of s_0, c_f and the coefficient a. The results of this calculation are given in Tables 1—3 and represented diagrammatically in Fig. 1—5.

In all these diagrams the abscissa represents the molal concentration of the solvent employed, and the ordinate the solubility ratio s/s_0. The solubility s_0 being assumed to be negligeable in the case represented in fig. 1, this diagram shows only the influence of the a-value i. e. the influence of the valency. This influence is seen to be exceedingly marked. Whilst the solubility of an uni-univalent salt is only moderately affected by addition of a 0.1 molal solvent, the effect in the case of a bi-bivalent

Table 1. The solubility ratio s/s_0 for $a = \frac{1}{3}$.
Heteroionic solvent.

c	$s_0 = 0$	10^{-4}	10^{-3}	10^{-2}	10^{-1}
0.001	1.080	1.045	1.020	1.006	1.001
0.01	1.180	1.139	1.100	1.045	1.013
0.02	1.232	1.189	1.146	1.078	1.025
0.05	1.327	1.281	1.232	1.148	1.058
0.1	1.428	1.378	1.325	1.228	1.105

Table 2. The solubility ratio s/s_0 for $a = \frac{1}{2}$.
Heteroionic solvent.

c	$s_0 = 0$	10^{-4}	10^{-3}	10^{-2}	10^{-1}
0.001	1.359	1.193	1.081	1.027	1.009
0.01	1.937	1.687	1.471	1.223	1.084
0.02	2.301	2.001	1.733	1.395	1.170
0.05	3.099	2.693	2.319	1.808	1.405
0.1	4.157	3.611	3.104	2.384	1.784

Table 3. The solubility ratio s/s_0 for $a = 3$. Heteroionic solvent.

c	$s_0 = 0$	10^{-4}
0.001	1.995	1.496
0.01	4.43	3.27
0.02	6.52	4.80
0.05	12.74	9.40
0.1	24.69	18.27

$s_0 =$	0.001		0.002		0.005
c		c		c	
0.0008	1.196	.....	....	.....	....
0.00337	1.633	0.00227	1.365	.....	....
0.00778	2.22	0.00629	1.856	0.00320	1.360
0.01673	3.27	0.01454	2.731	0.00999	2.002
0.04361	6.37	0.03933	5.337	0.03045	3.910
0.08762	12.38	0.07932	10.34	0.0621	7.578

$s_0 =$	0.01		0.02		0.05
c		c		c	
0.0053	1.47	.....	...	.....	...
0.0107	1.93	0.0038	1.31	.....	...
0.0160	2.40	0.0074	1.63	.....	...
0.0213	2.87	0.0110	1.95	.....	...
0.0267	3.37	0.0142	2.29	0.0015	1.17
0.0310	3.90	0.0172	2.64	0.0025	1.35
0.0358	4.42	0.0198	3.01	0.0030	1.54
0.0443	5.57	0.0243	3.79	0.0031	1.94

one is very considerable, and in the case of a tri-trivalent one enormous.

The influence of the solubility s_0 is shown in fig. 2 for uni-univalent salts. The smaller the solubility the more pronounced, obviously, is the effect of the added salt. The

same is the case by consideration of bi-bivalent salts. By the greater solubilities the curves approach straight lines.

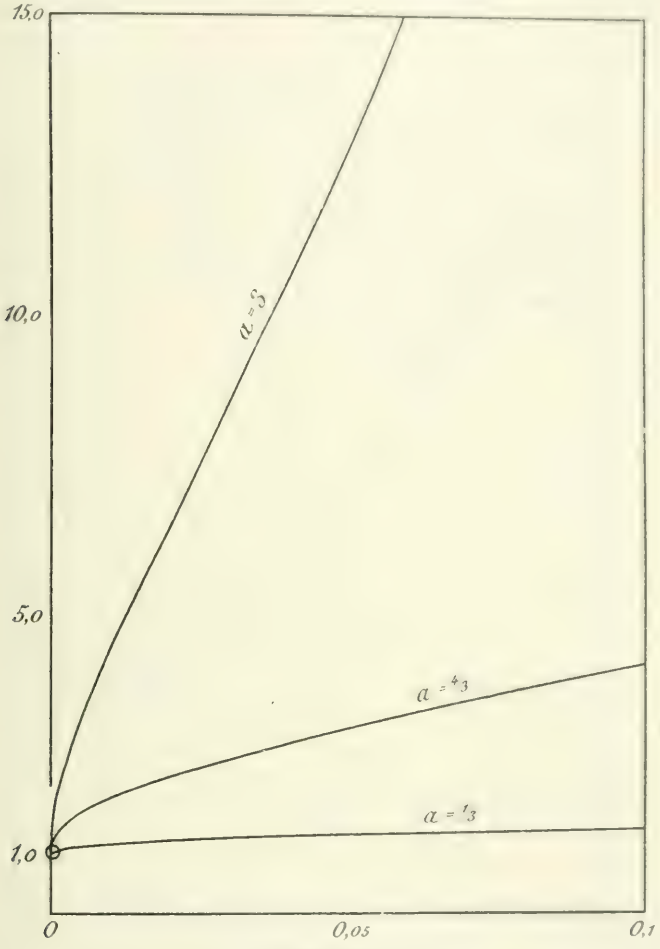


Fig. 12.

while the curvature is very marked in the case of slight solubilities

For tri-trivalents salts the results are plotted in fig. 3—5. It will be observed that the influence of s_0 on the shape

and the slope of the curves is similar to that in the foregoing cases only until the solubility $s_0 = 0.02$ molal. By

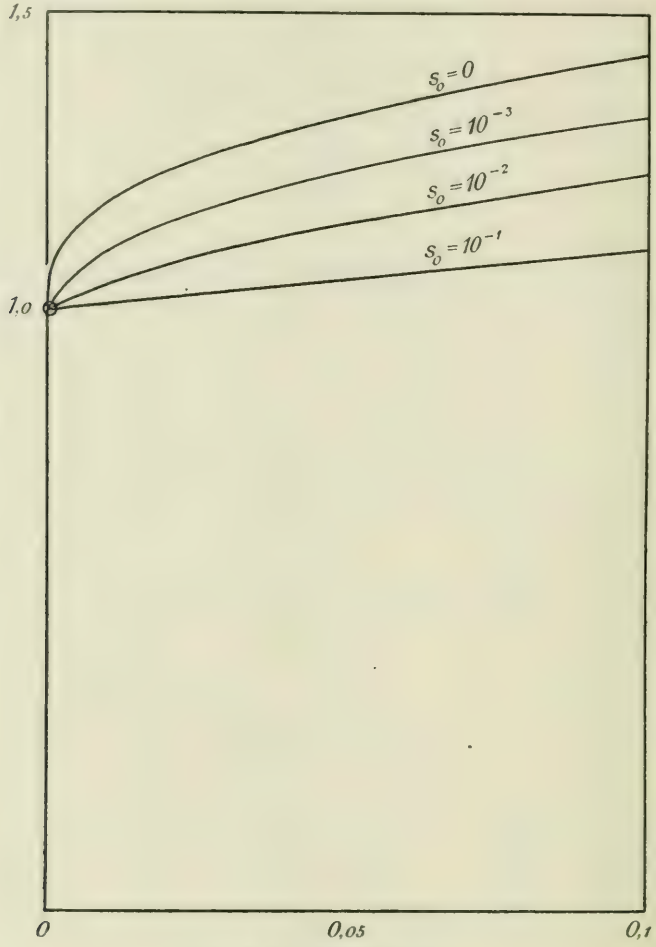


Fig. 2. Uni-univalent salts.

this concentration the effect exhibits a minimum, and by higher solubilities the curves rise again showing a curvature in the opposite direction of that otherwise found. In this respect the behaviour of tri-trivalent salts diverges

widely from that of the salts of lower types as will especially be evident on considering Fig. 5. The significance of the dotted line in this figure and in Fig. 3 will be mentioned later on.

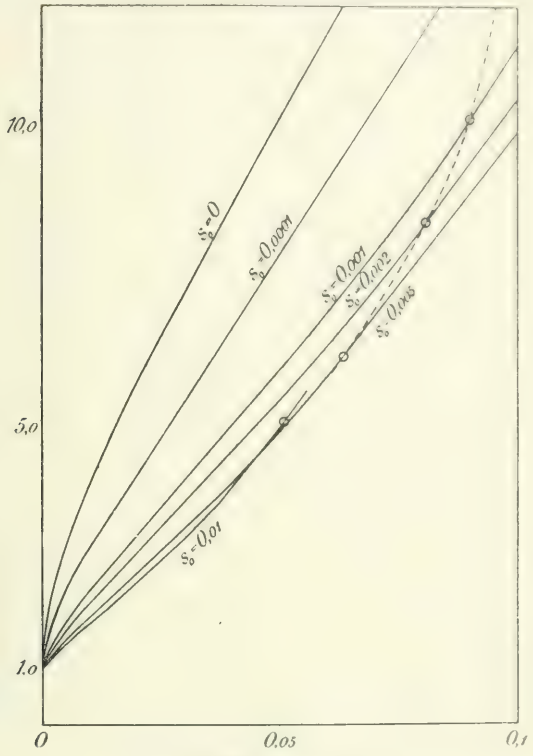


Fig. 3. Tri-trivalent salts.

In a simpler but somewhat less perspicuous way the calculations after (30) can be represented by employing the total concentration c_t instead of the concentration c of the solvent as abscissa. For this purpose we first transform equation (30) into the following

$$\log \frac{s_1}{s_2} = a\left(\sqrt[3]{c_{t_1}} - \sqrt[3]{c_{t_2}}\right) \qquad (31)$$

s_1 and s_2 representing the solubilities at the total concentrations c_{t_1} and c_{t_2} respectively. Putting here $c_{t2} = 0.1$ we obtain the figures given in Table 4 and represented in the graph fig. 6.

For each value of a only one single curve irrespective of the solubility s_0 is obtained. The results are conse-

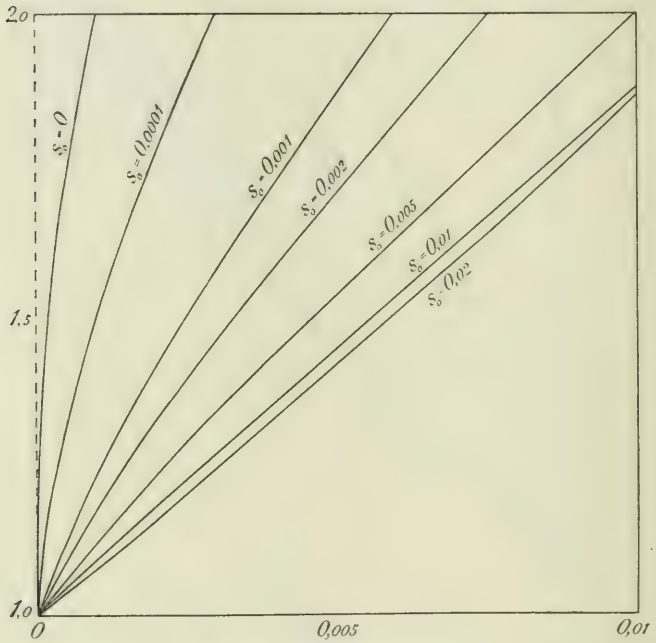


Fig. 4. Tri-trivalent salts.

quently more completely given by this diagram than by the foregoing. On the other hand it is more difficult to grasp the significance of s_0 by this method.

Having now shovn the general results furnished by the equations (30) and (31) in the case of heteroionic solvents we shall turn to their application to some definite examples. In spite of the great amount of experimental work

carried out on solubilities partly referred to in the introduction, only a few data are available for the purpose before us. A series of investigations have therefore been initiated in this laboratory for supplying this lack, and an ample experimental material has already been procured, chiefly covering salts of the metal ammonia type.

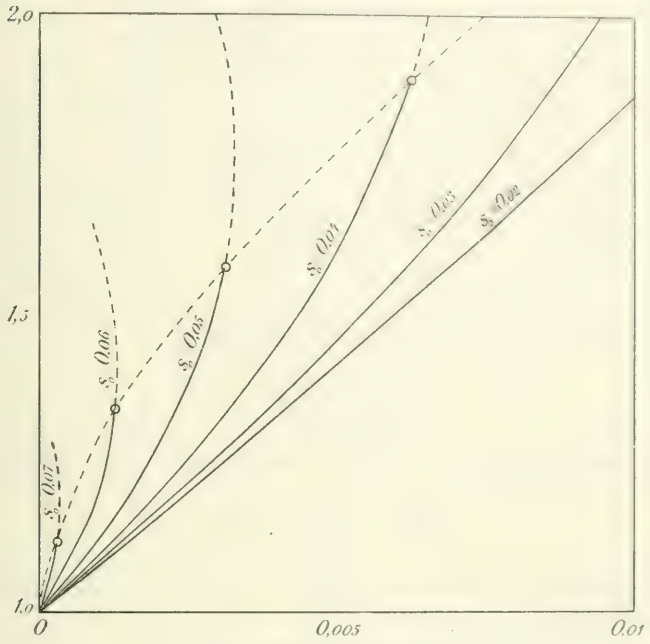


Fig. 5. Tri-trivalent salts.

From this series, of which a complete account will be rendered in subsequent articles, a few data may be stated in Tab. 5—9. The metal ammonia salts here in question are the following:

Dichlorotetrammine cobalt rhodanide or praseo rhodanide

$$\left[Co\begin{matrix}Cl_2\\(NH_3)_4\end{matrix}\right]Rh.$$

Table 4. The solubility ratio $s/s_{0,1}$ for different values of a Heteroionic solvent.

$c+s$	$a=\frac{1}{3}$	$a=\frac{1}{2}$	$a=3$
0.000	0.7003	0.2405	0.0405
0.001	7561	3269	808
0.002	7714	3541	967
0.005	7984	4065	1320
0.01	8262	4660	1795
0.02	8625	5535	2642
0.05	9291	7452	5159
0.08	.....	.....	7940
0.1	1.0000	1.0000	1.0000

Trans dinitrotetrammine cobalt nitrate or croceo nitrate

$$\left[Co\begin{matrix}(NO_2)_2\\(NH_3)_4\end{matrix}\right]NO_3.$$

Oxalotetrammine cobalt perchlorate

$$\left[Co\begin{matrix}C_2O_4\\(NH_3)_4\end{matrix}\right]ClO_4.$$

Nitropentammine cobalt chromate or xantho chromate

$$\left[Co\begin{matrix}NO_2\\(NH_3)_5\end{matrix}\right]CrO_4.$$

Further the data of Noyes and Bray[1] on thalleous chloride are shown.

Under $\left(\frac{s}{s_0}\right)_{\text{theor.}}$ are presented the solubility ratios calculated by means of (30), introducing for a a value such as to make the calculated and the experimental values of s/s_0 agree most closely. It is seen from the tables, firstly, that agreement can be brought about if we introduce one single a value for any system, secondly, that these a values for the various systems of uni-univalent salts vary only

[1] l. c.

slightly in the neighbourhood of $^1/_3$. In this respect full concordance exists with the requirements of the freezing-point formula. In the case of the bi-bivalent Xanthochromate the a value 1.80 must be adopted to represent the experimental data. This value also corresponds to the

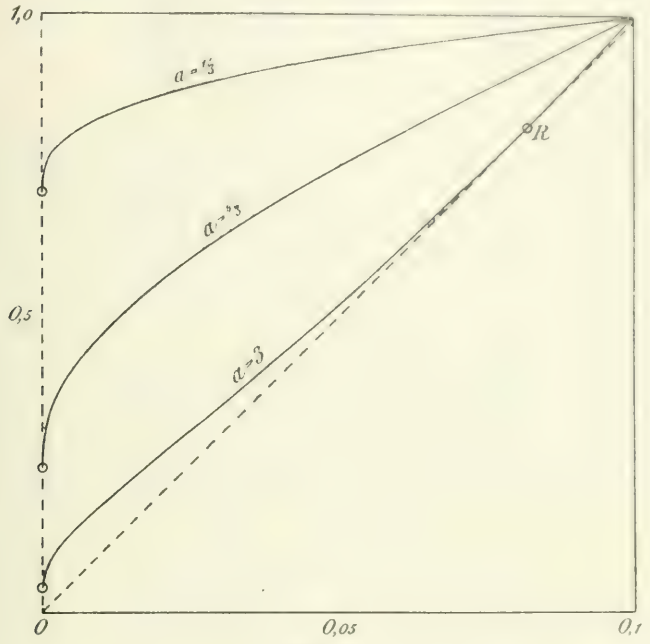


Fig. 6.

freezing-point measurements and is approximately 4 times as great as in the case of uni-univalent salts.

For tri-tri-valent salts no experiments have been carried out with tri-tri-valent solvents. The results obtained by employing other solvents, however, show the rise in solubility here to be extremely marked, highly exceeding the effect observed in the case of lower valency. This agrees with the corresponding higher values for a, which may be used in this case.

Table 5. The solubility of praseo cobalt rhodanide in natrium chlorate solutions at $t = 0°$. $a = 0.30$.

c	s	(s/s_0) theor.	(s/s_0) exp.
0.00	0.00289	1.000	1.000
0.01	306	1.066	1.059
0.03	327	1.132	1.132
0.05	339	1.175	1.173
0.1	360	1.254	1.246

Table 6. The solubility of croceo cobalt nitrate (β-modification) in kalium formiate solutions at $t = 0°$. $a = 0.31$.

c	s	(s/s_0) theor.	(s/s_0) exp.
0.00	0.00494	1.000	1.000
0.02	536	1.092	1.086
0.05	5745	1.163	1.164
0.1	621	1.242	1.258

Table 7. The solubility of oxalotetrammine cobalt perchlorate in natrium formiate solutions at $t = 0°$. $a = 0.33$.

c	s	(s/s_0) theor.	(s/s_0) exp.
0.00	0.00508	1.000	1.000
0.05	597	1.174	1.174
0.10	640	1.258	1.260

Table 8. The solubility of thalleous chloride in kalium nitrate solutions at $t = 25°$. $a = 0.355$.

c	s	(s/s_0) theor.	(s/s_0) exp.
0.00	0.01607	1.000	1.000
0.02	1716	1.068	1.068
0.05	1826	1.136	1.136
0.1	1961	1.217	1.220

Table 9. The solubility of xantho cobalt chromate in magnesium sulphate solutions at $t = 0°$. $a = 1.80$.

c	s	(s/s_0) theor.	(s/s_0) exp.
0.00	0.000258	1.00	1.00
0.02	620	2.39	2.40
0.05	908	3.57	3.52
0.1	1237	5.3	4.8

As far as illustrated by these facts the established formulae (30) and (31) may thus be looked upon as able to account for the shape of the solubility curves by varying valency of the ions by employment of heteroionic solvents. As will be observed on closer inquiry into the available experiments, the formulae hold generally only approximately in concordance with the individual behaviour which the various salts exhibit in solution, as known also from the individuality in the freezing-point curves.

5. Homoionic Solvents.

In the case of homoionic solvents the solubility phenomena are much more multitudinous than when heteroionic solvents are employed. The correctness of the point of view advanced for the treatment of solubility equilibria, will therefore be subjected to a more severe test by the employment of solvents of this kind.

From equation (27)

$$\log (c' c'') = 2a \sqrt[3]{c_l} + \text{const.}$$

we obtain by putting $c' = s$ and $c'' = c_l$ and elimination of the constant

$$\log \frac{s\, c_t}{s_0^{\,2}} = 2a\left(\sqrt[3]{c_t} - \sqrt[3]{s_0}\right) \qquad (32)$$

or

$$\log \frac{s_1\, c_{t_1}}{s_2\, c_{t_2}} = 2a\left(\sqrt[3]{c_{t_1}} - \sqrt[3]{c_{t_2}}\right) \qquad (33)$$

where the terms applied have the same meaning as above, and especially the product $s\, c_t$ indicates the stoichiometric solubility product. By application of these formulae to the various types of salts the corresponding values of a, given p. 14, must be introduced.

We shall first consider the effect of the addition of an homoionic solvent qualitatively. To that end equation (32) is differentiated. Hence we obtain

$$\frac{d\ln s}{d\ln c_t} = \frac{a}{0{,}6515}\sqrt[3]{c_t} - 1. \qquad (34)$$

If c_t be sufficiently small, the first member of the right-hand expression will vanish, and we can write

$$d\ln s = -d\ln c_t.$$

In this case the solubility product is constant. The solubility must therefore always decrease by increasing concentration of the homoionic solvent if the solubility in pure water be sufficiently small, and the solvent added in sufficiently slight concentration.

From (34) we draw the general inference that the solubility curve must exhibit a minimum when

$$\sqrt[3]{c_t} = \frac{0{,}6515}{a} \qquad (35)$$

showing by smaller concentrations a fall and by higher concentrations a rise by rise in concentration of the solvent. This minimum lies at very different concentrations according to the different values of a. Putting for a the

values adopted above for the various types of salts we obtain the following figures for the minimum concentration.

Table 10.

type of salt	a	$c_{min.}$
uni-uni-valent ...	$^1/_2$	7.5
bi-bi-valent	$^4/_3$	0.12
tri-tri-valent	3	0.01

In the case of uni-univalent salts the concentration calculated in this way lies far beyond the domain of the established formulae. For salts of higher types, however, the calculated $c_{min.}$ are sufficiently small to render the results reliable. We may therefore expect that a minimum in the solubility curve may occur for bi-bivalent and tri-trivalent salts at the total concentration in the neighbourhood of 0.1 and 0.01 molal respectively.

If the solubility in pure water of a bi-bivalent or a tri-trivalent salt exceeds these minimum values 0.1 and 0.01 molal respectively, the solubility curve must obviously start rising contrarily to the ordinary conception of the influence of a homoionic solvent.

We shall now turn to the numerical calculation of the effect of a homoionic solvent. It is first to be noted that formula (31) represented in fig. 6 applies very well also in the case of homoionic solvents when sc_t/s_0^2 is inserted for s^2/s_0^2. In fig. 6 the curves then represent the square root of the solubility product instead of the solubility itself. A more perspicuous picture of the correlations in question will however be obtained when, as in the foregoing section, the interdependence between the solubility ratio s/s_0 and the concentration of the solvent is represented.

The results of these calculations according to equation (32) are given in the following tables 11—13 and the accompanying diagrams.

Table 11. The solubility ratio s/s_0 for $a = \frac{1}{3}$. Homoionic solvent.

c	s/s_0	c	s/s_0	c	s/s_0
0.0001	0.108	0.01	0.119	0.01455	0.545
0.01	0.01296	0.02	0.0651	0.0475	0.253
0.02	0.00706	0.05	0.0302	0.0999	0.146
0.05	0.00328	0.1	0.0175	...	...
0.1	0.00190	...	...	...	...
$s_0 = 10^{-4}$		$s_0 = 10^{-3}$		$s_0 = 10^{-2}$	

Table 12. The solubility ratio s/s_0 for $a = \frac{4}{3}$. Homoionic solvent.

c	s/s_0	c	s/s_0	c	s/s_0
0.001	0.139	0.0098	0.197	0.01295	0.705
0.01	0.0282	0.0199	0.143	0.0445	0.512
0.02	0.0199	0.0499	0.104	0.0954	0.461
0.05	0.0144	0.0999	0.094	...	...
0.1	0.0129	...	...	...	...
$s_0 = 10^{-4}$		$s_0 = 10^{-3}$		$s_0 = 10^{-2}$	

Table 13. The solubility ratio s/s_0 for $a = 3$. Homoionic solvent.

c	s/s_0	c	s/s_0	c	s/s_0
0.00098	0.210	0.00951	0.493	0.0092	1.084
0.01	0.1033	0.0195	0.534	0.0334	1.656
0.02	0.1120	0.0492	0.816	0.0689	3.11
0.05	0.1710	0.0985	1.532	...	...
0.1	0.320	...	...	...	...
$s_0 = 10^{-4}$		$s_0 = 10^{-3}$		$s_0 = 10^{-2}$	

We observe from these that the equation (32) underlying the calculations is able to account for the wide diversities which, according to the available data in this field, are characteristic of solubilities in homoionic solvents. By small a-values or [uni-uni-valent salts the solubility

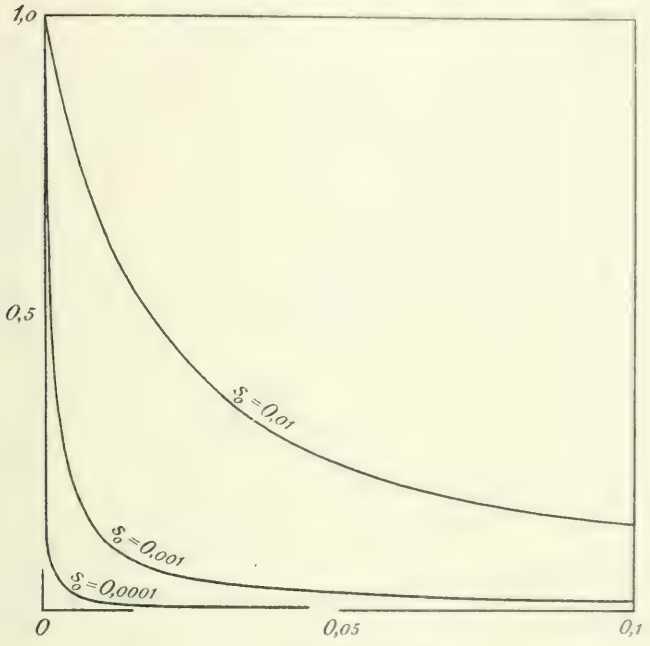


Fig. 7. Uni-univalent salts

exhibits a pronounced and steady decrease by increasing concentration of the added salt. By medium values of a or bi-bi-valent salts the solubility is likewise diminished, but the fall is less marked than in the foregoing case, and the curves gradually approach a horizontal line. To a large value of a or tri-tri-valent salts, finally, curves correspond showing pronounced minima and, by higher concentration, a rapid rise in solubility. In all cases the shape of the

curves is markedly influenced by the value of the solubility s_0 in pure water. This value being sufficiently high we find in the case of a tri-tri-valent salt, as mentioned above, that the decreasing portion of the curve will disappear and the curve start rising.

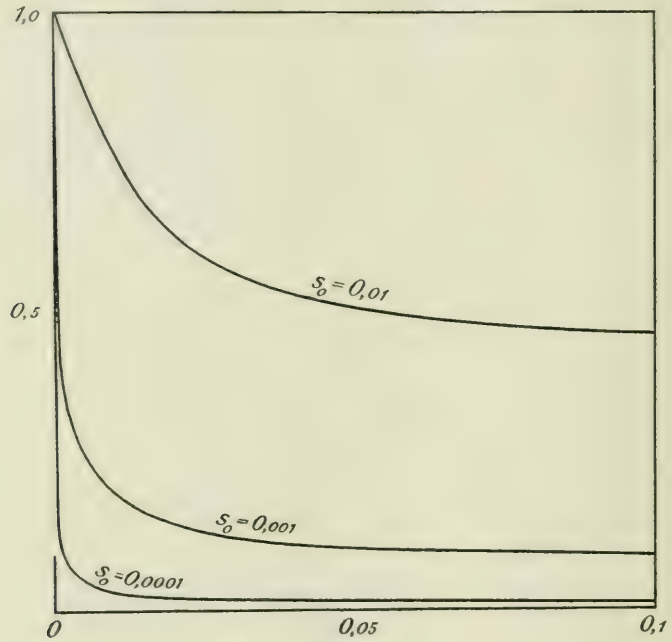


Fig. 8. Bi-bivalent salts.

A few of the results which have been furnished by the examination of metal ammonia salts dissolved in homoionic solvents may be stated here for a closer illustration of the applicability of the theory. The salts in question are the following:

Dichlorotetrammine cobalt bromide or praseo bromide

$$\left[Co \begin{matrix} Cl_2 \\ (NH_3)_4 \end{matrix} \right] Br.$$

Trans dinitrotetrammine cobalt chloride or croceo chloride

$$\left[Co\begin{matrix}(NO_2)_2\\(NH_3)_4\end{matrix}\right]Cl.$$

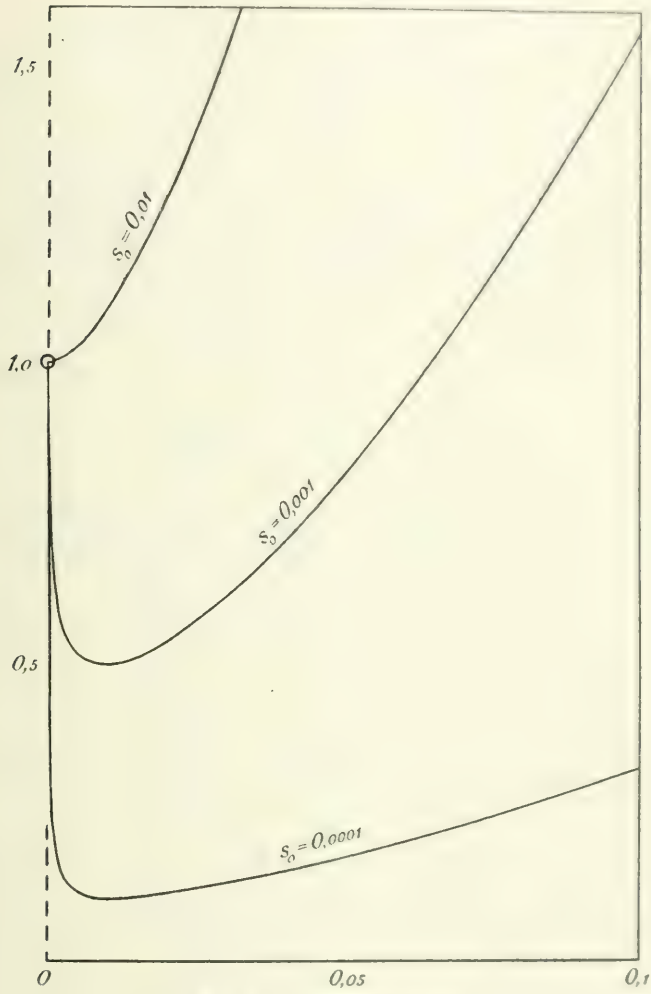


Fig. 9. Tri-trivalent salts.

Isorhodanopentammine cobalt sulphate

$$\left[Co\begin{matrix}Rh(SCN)_3\\(NH_3)_5\end{matrix}\right]SO_4.$$

Table 14. The solubility of praseo bromide in kalium bromide solutions at $t = 0°$. $a = 0.355$.

c	$s_{exp.}$	$s_{theor.}$
0.00	0.00635	—
0.02	0.00212	0.00212
0.05	0.00107	0.00107
0.1	0.000635	0.000634

Table 15. The solubility of croceo chloride in kalium chloride solutions at $t = 0°$. $a = 0.38$.

c	$s_{exp.}$	$s_{theor.}$
0.00	0.02216	—
0.02	0.01515	0.01516
0.05	0.01005	0.00991
0.1	0.00640	0.00647

Table 16. The solubility of isorhodanopentammine cobalt sulphate in magnesium sulphate solutions at $t = 0°$. $a = 1.73$.

c	$s_{exp.}$	$s_{theor.}$
0.00	0.01177	—
0.02	0.01005	0.0096
0.05	0.00980	0.0099
0.1	0.01034	0.0116

Table 17. The solubility of thallous chloride in kalium chloride solutions at $t = 25°$. $a = 0.40$.

c	$s_{exp.}$	$s_{theor.}$
0.00	0.01607	—
0.025	0.00869	0.00873
0.05	0.00590	0.00587
0.1	0.00396	0.00391

The data for these salts and the data of Bray and Winninghoff[1] for thalleous chloride are given in Tables 14—17.

The agreement between the calculated values and those found experimentally in all cases proves satisfactory. The values of a to be chosen for producing this agreement are quite concordant with the a values found by employment of heteroionic solvents and by the freezing point measurements. We observe especially that the a values for the bi-bivalent salt must again be chosen nearly 4 times as great as for uni-univalent ones. The appearance of a minimum in the curve for the bi-bivalent salt is likewise in accordance with the theory as mentioned above.

6. The Liquid-Liquid Equilibrium.

Some interesting phenomena whose occurrence may be suspected from the established formulae, say (30), in the case of polyvalent ions, may be mentioned now. As seen from the diagram fig. 6 the curvature of the curves for small values of a has the same sign through the whole range of concentrations. As a becomes greater the course of the curves becomes more rectilinear at the higher concentrations and, finally, when a sufficiently great value of a is reached, an inflexion point appears in the curve. Beyond this point the curvature is inverted, and at a certain point (R) the tangent of the curve may pass through the origin.

This case is more clearly represented in fig. 10 for addition of a homoionic salt. The solubility s is plotted against the total concentration c_t. The point R corresponds to the mentioned point R in fig. 6. We have here

$$\frac{ds}{dc_t} = \frac{s}{c_t}$$

[1] Journ. Am. Chem. Soc. 33, 1663 (1911).

or $$d\ln s = d\ln c_t = d\ln c. \tag{36}$$

Introducing here $s = \frac{\xi_1}{f}$ and $c = \frac{\xi_2}{f}$, ξ_1 and ξ_2 denoting respectively the activities of the saturating salt s_1 and the salt s_2 serving as solvent, we obtain

$$d\ln \xi_1 = d\ln \xi_2 \tag{37}$$

or, since ξ_1 = constant on account of the solution being saturated with s_1

$$d\xi_2 = 0. \tag{38}$$

At concentrations corresponding to the point of contact R the activity of s_2 will thus remain constant by rising concentration. While in all stable solutions addition of any substance must cause the potential or activity of this substance to rise we have here arrived at a critical point for the stability of the solution. According to the theory of miscibility of liquids a separation in two liquid layers must take place before this critical concentration is attained. The situation of this point is determined by (36). Combining this with (28) we obtain

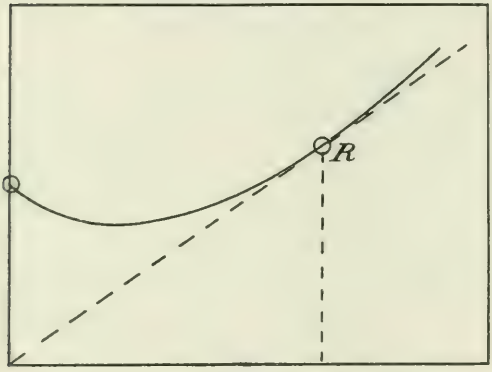


Fig. 10.

$$\sqrt[3]{c_R} = \frac{1.303}{a} \tag{39}$$

where c_R indicates the mentioned critical concentration.

According to equation (18) the point R is also characterized by

$$d\eta = 0$$

which indicates that the osmotic concentration here shows a maximum. Since this will be the case also in pure salt solutions, the phenomenon of separation here in question must occur in pure salt solutions too.

The same result is easily found by applying formula (26)

$$\log f = -a\sqrt[3]{c}.$$

Putting here $f = \xi/c$ we have

$$\log \xi = \log c - a\sqrt[3]{c}$$

or, since $d \log \xi = 0$ by $c = c_R$

$$\sqrt[3]{c_R} = \frac{1.303}{a}$$

corresponding to (39).

As a result of these computations we see that the activity of a dissolved salt, increasing by increasing concentration in dilute solution, by higher concentrations will reach a maximum and thereafter decrease. These computations of course have a physical meaning only if the critical point R is reached by solutions which are sufficiently dilute for the applicability of the involved formulae. Introducing in (39) the values of a which are used above as fitted to represent the behaviour of the various types of salts, we shall calculate the following values of c_R.

Table 18.

Type of salt	a	c_R
uni-univalent	1/3	59
bi-bivalent	4/3	0.93
tri-trivalent	3	0.082

Only in the case of tri-tri-valent salts, as seen from the figures, the calculation of c_R affords some trustworthy

result, as only in this case the magnitude of c_R calculated is small enough for rendering the applied formulae to a certain degree valid. Even the concentration 0.08 computed for such salts may be regarded as a rough estimation, on account of the deviation from the used formulae increasing rapidly with increasing valency of the ions. We are justified, however, in expecting the separation phenomenon in the case of polyvalent ions at least at fairly small concentrations.

Similarly, the separating layer is expected to contain a small amount of the solvent. It may naturally be regarded as a salt in liquid state saturated with water under the prevailing conditions. If the melting point of the salt be very low the liquid-liquid system in certain cases may be a stable one. Usually, however, the equilibrium is able to exist only owing to the insufficient velocity of crystallization, the tendency of the concentrated layer to separate the solid phase thereby being obscured.

According to this view the solubility of salts of higher types may remain far below the solubility of the liquid layer[1] or a forteriori far below the critical concentration c_R. In fact the solubility determinations with such salts, carried out in connection with the mentioned investigation on metal ammonia salts, clearly confirms this conclusion, as in all cases solubilities not surpassing 0.001 molal was found.

Regarding the separation of a liquid layer from a salt solution, this phenomenon has been realized in a number of cases by admixture of solutions of polyvalent ions with different sign. The separation is in all cases studied by mixing two samples of the solutions concerned under the

[1] See J. N. Brønsted: Zeitschrift f. physik. Chemie 56, 683 (1906).

microscope. Usually the separation occurs instantaneously forming clouds of minute particles gradually joining to greater liquid drops. The change of the liquid salt into the solid phase usually takes place rapidly. Notably in the case of the tri-valent triethylene-diamine cobalt ion the liquid system, however, may be fairly stable, and the procedure of the growth of the drops and their ultimate crystallization furnish a proper illustration of the nature of the phenomenon.

It remains to be said that the phenomenon of the appearance of a salt in liquid state by precipitation is known from other fields of research. The interest of the phenomenon, and its importance for our conception of salt solutions, however, makes a closer theoretical and experimental investigation desirable. We shall there have the opportunity of considering the experimental results from this laboratory and from other quarters more closely.

In Fig. 3 and 5 the critical points on the curves are indicated by circles. The dotted line drawn through these points forms the limit of stable solutions. All parts of the curves beyond this limit are dotted in the figures in order to show the instable or hypothetical character.

By employment of homoionic solvents a limiting curve of similar meaning can be drawn. In all cases the critical points can be determined as points of contact of the tangent passing through the origin.

7. Theoretical Considerations.

Through the results obtained in the foregoing sections we are justified in concluding that the solubility phenomena of salts in solutions of other salts with or without an ion in common may be accounted for on a thermo-

dynamic base by the simple assumption of the equality of the activity coefficients for salts or ions of the same type in a mixed salt solution.

In particular cases this assumption may prove to hold very exactly, f. inst. in mixed solutions of kalium chloride and kalium bromide of slight concentration. Solutions of such a kind will behave in a markedly simple manner by constant total concentration, and may suitably be termed ideal salt solutions. Apart from such solutions, the assumption of the equality of the activity coefficients, as already emphasized in the foregoing, pretends to be only an approximation. It is important to note, however, that the activity coefficients of mixed salts are not able to vary independently but, as mentioned in section 2 of this article, are subjected to thermodynamical correlations. In a following paper we shall deal with these correlations more thoroughly.

Like our chief assumption concerning the activity coefficient, the employed cube root expression for $\log f$ is also an approximation. It is clear, therefore, that all the formulae derived in the preceeding chapters represent only the general shape of the solubility curves, but are not able to explain the peculiar behaviour of a selected system. In order to ascertain whether the hypothesis advanced may, properly modified, be able to form the base of a more minute theory of solubility, it must preferably be looked upon in the light of the theory of the complete electrolytic dissociation supported by Bjerrum[1] and Milner[2].

In a recent paper[3] I have shown that this theory may

[1] Proc. of the 7. intern. Congress of appl. Chemistry London 1909. Sec. X. Zeitschr. f. Elektrochemie 24, 321 (1918).

[2] Phil. Mag. 6, **23**, 553 (1912), 6, **25**, 742 (1913).

[3] Meddelanden fran K. Vetenskabsakademiens Nobelinstitut **5** Nr. 25.

be applied advantageously in the case of mixed salt solutions to explain the decrease in activity exhibited by a dissolved salt on addition of another salt. From this point of view the raised solubility of salts in heteroionic solvents follows simply from the diminution of the activity coefficient of the ions, no assumption as to the formation of undissociated molecules being involved. In the case of a homoionic addition the solubility product for the same reasons must rise, producing either a rise or a fall in the solubility itself, according to the type of the added salt and the prevailing conditions of the system as a whole.

According to the definite character of the activity in any given solution, the stoichiometric activity coefficient defined in chapter 2 and reckoned with through this paper must obviously be equal to the activity coefficient arrived at by the suggestion of complete dissociation, since in this case the actual concentration of the ion is equal to its stoichiometric concentration.

If the calculations carried through by MILNER, in order to determine the influence of the electric charges of the ions, held good for real salts, the activity should be the same in all cases for the same type of salt, and the above approximate hypothesis would then be a rigid law. The deviations manifested by salt solutions, when examined with regard to freezing-point or solubility, proves, however, that the ions possess a markedly separate or individual behaviour which is appreciable already at fairly small concentrations. From the usual point of view this individuality should be ascribed to diversities in the degree of dissociation. On the base of the theory of complete dissociation it becomes necessary to attribute to any separate ion a secondary power beside the primary MILNER effect

able to exert a peculiar action on ions of the same or of another type. The necessity of introducing such a secondary effect becomes quite natural when we bear in mind that the computations of Milner rest on the presupposition of ideal conditions in the solutions, the ions being considered as charged mass points, and no regard being paid to the contingency of hydration of the ions. When the ions are considered as particles of finite sizes, and appointed with electrical charges in finite distance from one another, and moreover often combined with a large number of water molecules, they must behave otherwise than the ideal ions which Milner deals with. It is easily understood, therefore, that the theory of Milner is not able to give an exact picture of the behaviour of salt solutions.

Considering now the laws of the strong electrolytes as being governed merely by these primary and secondary forces, the interpretation of several solubility phenomena which, according to the ordinary view of such solutions, require particular assumptions for their explanation will appear specially simple. To explain the increasing solubility, which is exhibited by sparingly soluble salts on addition of a homoionic solvent in fairly considerable concentration, we thus commonly suspect the formation of "complex" ions in the solution to take place. Of course in many cases this opinion may be true. The solubility of silver cyanide in kalium cyanide f. inst., is undoubtedly due to the formation of a complex silver cyanide ion. The rise in solubility produced in this way is no exceptional phenomenon, however, but generally met with, as mentioned above, when the coefficient a has reached a certain value. From the theory advanced the explanation is given by the double action of the added salt in raising the activity

on account of the increasing concentration of the ion concerned, and in causing it to fall on account of the rising electric forces operating between the ions, the last mentioned effect exceeding the first in the case of raised solubility. In a solution of hexamine cobalt sulphate f. inst., we have no reason for believing the ions tending to form complexes. Nevertheless the solubility of this salt increases rapidly by increasing concentration of a natrium sulphate solution serving as solvent. In this as in similar cases the explanation follows as a simple consequence of the electric action which must be associated with the ions according to our theory.

In this connection we may mention a phenomenon to which attention is called by NOYES and BRAY[1]. It is stated by these observers that, though in all cases of moderately soluble uni-bivalent salts the solubility is reduced in at least rough conformity with the requirements of the solubility product principle when a salt is added having a univalent ion in common, the decrease of the solubility caused by the addition of an equivalent quantity of a salt with a common bivalent ion does not exceed a few per cent, and the addition of a larger quantity causes the solubility to increase slightly. In other words — the effect on the solubility in the last mentioned case is on the whole very slight. HARKINS[2] emphasizes this phenomenon as having no resemblance to what would be expected from an application of the solubility product principle.

From our point of view, however, a distinction between the two cases does not exist. Consider a uni-bivalent salt, f. inst. lead chloride. Let s_1 and s_2 be the solubility in

[1] l. c.

[2] Journ. Am. Chem. Soc. 33, 1807 (1911).

equivalents in a solution of kalium chloride of the concentration c_1 and in a solution of lead nitrate of the equivalent concentration c_2. Then we have

$$\frac{s_1}{2}(s_1 + c_1)^2 f_1 f_1'^2 = \frac{s_2 + c_2}{2} s_2^2 f_2 f_2'^2$$

where f_1, f_1' and f_2, f_2' indicate the activity coefficient of the lead ion and the chlorine ion in the first and second solution respectively. Assume the total equivalent concentration to be equal ($= c_t$) in the two cases, and the activity coefficients of the two ions to depend only on the total concentration. The above equation then simplifies to

$$s_1 c_t = s_2^2. \tag{40}$$

This equation shows that a rapid fall in solubility by increasing concentration of an added salt having a univalent ion in common with the saturating salt is well compatible with the solubility being unaltered by addition of a salt with a common bivalent ion. As in dilute solutions the gas laws approach validity, s_1 will decrease more rapidly in this region than corresponding to $s_1 c_t = k$ and s_2 consequently will not be constant but decrease slightly. By increasing concentration the fall in s_1 may gradually diminish, corresponding to s_2 passing through a minimum and thereafter increasing. The effect observed is thus no abnormal one but contrarily in full concordance with what would be expected from the standpoint adduced.

The significance of the valency of the ions is accounted for by assigning different values to the coefficient a in the employed formulae according to the type of the salts. The relationship between valency and activity coefficient is suggested by Bjerrum[1] and explained through the attracting

[1] Zeitschr. f. Elektrochemie 24, 321 (1918).

forces operating between the ions, these forces changing proportionately to the product of the electric charges. No formal derivation of this dependency, however, is given and at the present moment it is not quite decided whether the molal or the equivalent concentration applies most profitably in the formulae. At any rate, the rapid rise in the electric forces by increasing valency manifests itself very distinctly by the solubility measurements.

As shown in the foregoing there will be a fall in the activity by rising concentration if the coefficient a be sufficiently great, that is to say, in the case of tri-trivalent salts. This is easily understood when we consider the solution constituted by ions of high charges. As the concentration increases, the attracting forces between the ions will tend to diminish the activity contrarily to the effect otherwise prevailing in dilute systems, and at a certain point, therefore, the activity will pass through a maximum. This corresponds to the separation of another liquid layer in equilibrium with the solution, as mentioned above. This phenomenon is quite similar to the breach of continuity by compression of a gas beyond the saturation point, the cohesive forces between the molecules here replacing the electric forces operating between the ions. As the incomplete miscibility of liquids according to the theory of HILDEBRAND[1] is sensibly dependent upon the difference in internal pressure of the liquids concerned, we must suggest the supercooled liquid salts of high valency to be in possession of extremely high internal pressure, as exactly to be expected from the conception of the salts being, in solid and liquid state, also more or less completely dissociated. At any rate the assumption of a decreasing dissociation of the salts by

[1] Journ. Am. Chem. Soc. 38, 1452 (1916).

rising concentration may not be able to account for the separation phenomenon appearing in very dilute solutions in a similarly simple way.

Finally, we may consider briefly some other fields in which the theory supposedly may be advantageously applied. It is clear, namely, that the effect, which is here assumed to render the solubility of a salt increasing by addition of other salts, must exist also in the case of unsaturated solutions and may be able to alter the equilibrium occurring in homogeneous solutions as a whole.

Consider firstly the electrolytic dissociation of the solvent itself. In accordance with the above view an addition of salt to pure water must diminish the activity coefficient of the hydrogen- and the hydroxyle-ion. As on the other hand the potential of the water suffers no perceptible alteration on small additions of salt, a rise in the dissociation of the water must ensue. The "dissociation constant" is thus actually no constant, but increases with the concentration of the solute, say a uni-univalent salt, in the same way as the solubility product of a correspondingly soluble salt. If by k we denote the dissociation product in pure water, the value of the product of the two ions at varying concentration of an added uni-univalent salt,

Table 19. The dissociation product of water in salt solutions of the concentration c.

c	$c_{H+} \cdot c_{OH-}$
0	k
0.001	1.17 k
0.01	1.39 k
0.02	1.51 k
0.05	1.77 k
0.1	2.05 k

calculated by means of (27), may be expressed approximately by the following figures.

The alteration in the dissociation product is thus very considerable. In a 0.1 molal solution it is about twice the value in pure water. That such a change has not been observed hitherto is due to the fact that usually only the activity of the ions, and not their true concentrations, are available for the experiment. If the activity coefficient be altered to the same degree for the two ions, the effect must then be obscured[1].

As in the case of water, we shall find an effect of the same kind by other weak electrolytes. We may establish quite generally that any equilibrium in aqueous solution involving incompletely dissociating electrolytes may be influenced correspondingly by addition of salts. By consideration of such cases, however, the phenomenon known as the "salting out" effect must not be disregarded. This effect will cause the activity of a non-electrolyte, or of the undissociated portion of a weak elektrolyte, to rise by rising salt concentration. The solubility of a slightly soluble weak acid f. inst., will for these reasons exhibit a maximum at a certain salt concentration at which the MILNER effect is compensated by the "salting out" effect. Experiments of HOFFMANN a. LANGBECK[2] with weak acids have given results which are confirmatory to this view.

[1] If the $H+$ and the $OH-$ ions are hydrated to a different extent, rising concentration of an added substance must alter the dissociation equilibrium of water in favor of the anhydride of the most hydrated ion. This fact being appreciable in strong solutions makes it possible to explain a divergence of such solutions from the neutral state, as indicated by electrometric measurements. Possibly, the results of PALMAER and MELANDER (Zeitschr. f. Elektrochemie 21, 418 (1915)) are to be interpreted in this way.

[2] HOFFMANN a. LANGBECK, Zeitschr. f. physik. Chemie 51, 385 (1905).

For the treatment of solubility phenomena of the strong electrolytes the results obtained with non-electrolytes or weak electrolytes, however, are of wide consequence. Then it may be surmised that the "salting out" effect is not restricted to uncharged molecules alone but is present in the case of widely dissociating salts also. It is very probable that the "secondary effect" mentioned in the previous sections as being peculiar to the individual salts, respecting their dissolving power, may be found to be in close connection with the salting out effect. As an illustration may be noted that kalium hydroxide, showing the most pronounced salting out effect upon non-electrolytes, also reduces the solubility of salts much more markedly than other solvents examined in this respect[1]. This fact has caused A. E. Hill[2] to surmise without paying attention to the exceptional character of the employed substance that the solubility generally diminish on addition of heteroionic solvents.

Having now in the present paper advanced the general points of view from which the solubility equilibrium in salt solutions supposedly ought to be considered, we may in the following articles of this series turn to a more detailed treatment of the whole problem in connection with the publication of the remaining part of the experimental results.

[1] J. N. Brønsted, Meddelanden fr. K. Vetenskapsakademiens Nobelinstitut Bd. 5, Nr. 25 (1919).

[2] A. E. Hill, Journ. Amer. Chem. Soc. 32, 1186 (1910).

Royal Polytechnical Institute of Copenhagen.
May 1919.

Forelagt paa Mødet d. 11. April 1919.
Færdig fra Trykkeriet d. 10. December 1919.

CONTENTS

		Page
1.	Introduction	3
2.	Thermodynamic Functions Applicable to Solutions	6
3.	The Theory of Solubility of Salts in Salt Solutions	11
4.	Heteroionic Solvents	14
5.	Homoionic Solvents	25
6.	The Liquid-Liquid Equilibrium	33
7.	Theoretical Considerations	36

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **II,** 11.

DIE KAUSALE RELATIVITAETSFORDERUNG UND EINSTEINS RELATIVITAETSTHEORIE

VON

HELGE HOLST

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1919

I

Ein Hauptgrund für den starken Widerstand, welcher sich wider die Einstein'sche Relativitätstheorie erhoben hat, ist zweifellos darin zu suchen, dass die Aenderungen von den Dimensionen der Körper und der Dauer der physikalischen Prozesse, mit denen man in der Relativitätstheorie rechnen muss, wie es scheint, mit dem Ursachsgesetze unverträglich sind. Die Auffassung, dass es sich hier wirklich um einen unausgleichbaren Gegensatz handelt, wird nicht nur von der grossen Mehrzahl der Gegner gehegt; auch viele Anhänger der Theorie werden zugeben, dass sie nicht die gewöhnlichen kausalen Ansprüche erfüllt. Besonders stark und offen hat sich J. Petzoldt — ohne deswegen von Einsteins Lehre Abstand zu nehmen — in dieser Richtung in einer Abhandlung über gewisse Seiten der erkenntnistheoretischen Stellung der Relativitätstheorie geäussert (»Ber. der Deutschen Physikal. Ges.« Jahrg. 1918, S. 189); sich auf die erwähnten Aenderungen beziehend sagt er: »Damit wird nicht nur der alte Kraft-, sondern auch der Ursachsbegriff überhaupt über Bord geworfen«.

Es lässt sich denn auch nicht leugnen, dass die spezielle Relativitätstheorie, welche für die Stellungnahme der meisten Physiker entscheidend gewesen ist, für sich betrachtet wenigstens scheinbar an einem sehr auffallenden

Mangel an Kausalität leidet, und so lange nur dieser Anfang des imponierenden Gebäudes, das Einstein errichtet hat, bekannt war, mochte es den Anschein haben, dass man nur zwischen den beiden Alternativen zu wählen habe: entweder das alte Kausalitätsprinzip aufzugeben, oder das neue und überaus fruchtbare von Einstein aufgestellte Relativitätsprinzip zu verwerfen. Wenn man — wie einige Anhänger der Relativitätstheorie es tun — sagen würde, dass die Bewegung eines materiellen Systems relativ zum Beobachter Ursache zu Aenderungen der Länge von Massstäben und des Ganges von Uhren in dem Systeme ist, dann würde man jedenfalls das Wort *Ursache* in einem ganz anderen Sinne anwenden als in demjenigen, worin wir bisher den Begriff *einer physikalischen Ursache* auffassten. Denn was hier bei dem Beobachter für die Aenderungen in dem bewegten System bestimmend ist, ist doch eben nur die Eigenschaft, dass er Beobachter ist — also etwas Mentales — und durchaus nicht seine physikalischen Eigenschaften. Die Aenderungen sind ja nicht von seiner Masse, seiner Entfernung von dem Systeme oder dergleichen abhängig; seine Masse kann so gering, seine Entfernung so gross sein, dass er physikalisch als eine »*quantité négligeable*« angesehen werden muss. Die gewöhnliche kausale Auffassung lässt sich auch nicht dadurch retten, dass man die Aenderungen als blossen Schein — etwas, das nur für den betreffenden Beobachter existiert — betrachtet. Dass dies nicht angeht, wird besonders klar durch das bekannte Beispiel mit zwei Uhren, die sich zu Anfang beide an derselben Stelle in Ruhe relativ zueinander befinden und dann genau denselben Gang und Stand haben, aber, nachdem die eine Uhr eine Reise ausgeführt hat und in die ursprüngliche Ruhestellung zurückgekehrt ist, einen ver-

schiedenen Stand aufweisen. Hier tritt das Resultat des Vorganges in der verlaufenen Zeit als ein realer Unterschied zwischen den beiden Uhren hervor — ein Unterschied, der von jedem Beobachter erkannt werden kann — und es geht hieraus hervor, dass wenigstens die eine von den Uhren eine reale Veränderung erfahren hat. Auch die realen Einwirkungen, welche in Wirksamkeit treten müssen, um die Geschwindigkeit eines Körpers relativ zu einem von den berechtigten Koordinatensystemen der speziellen Relativitätstheorie zu verändern, vermögen keine befriedigende kausale Erklärung des auftretenden Unterschiedes zwischen den beiden Uhren zu liefern, wenn die verschiedenen berechtigten Koordinatensysteme als vollkommen gleichgestellt betrachtet werden sollen.[1]

[1] Dies wird aus folgenden Betrachtungen hervorgehen:

K sei ein »berechtigtes galileisches Koordinatensystem« d. i. ein solches, relativ zu welchem isolierte Körper sich gleichförmig und geradlinig bewegen, und K' ein anderes berechtigtes Koordinatensystem, das relativ zu K sich mit der Geschwindigkeit v im Sinne der positiven X-Achse (nach rechts) bewegt. Bei A befinden sich, ruhend in K, zwei Uhren U_1 og U_2 desselben Ganges und Standes. Jetzt wird U_2 durch eine äussere

Kraft nach rechts beschleunigt, bis sie in B die Geschwindigkeit v erreicht hat. Mit dieser Geschwindigkeit (folglich in K' ruhend) bewegt sie sich von B bis C. Eine äussere Kraft in der negativen Richtung setzt nun erst die Geschwindigkeit auf 0 herab (in D) und danach wieder herauf auf v nach links (in C). Sie geht mit dieser Geschwindigkeit zurück nach B, wonach eine äussere Kraft die Geschwindigkeit herabsetzt, bis sie in A wieder den Wert 0 erreicht hat; und man wird dann finden,

Viele Forscher haben jedoch versucht, die Annahme von Einsteins Lehre als gemeingültiges phänomenologisches Prinzip mit der kausalen Auffassung zu vereinbaren, indem sie am Weltäther als Erklärungsgrund festhielten und mit FitzGerald und H. A. Lorentz annahmen, dass durch die Bewegung relativ zum Aether eine Kontraktion in der Bewegungsrichtung hervorgebracht werde. Und nachdem Lorentz gezeigt hatte, dass eine sehr einfache Hypothese von der Natur der Kräfte zwischen den Teilen der Körper

dass U_2 jetzt relativ zu U_1 zurückgeblieben ist, dass sie aber den nämlichen Gang wie diese hat.

In den Akzelerationsperioden könnte man sich den Gang der Uhr auf verschiedene Weise geändert vorstellen:

1. Der Gang wird nur momentan geändert, d. h. er ist nur während der Akzelerationsperiode geändert. — Diese Annahme kann nicht richtig sein, da die schliessliche Aenderung im Stande der Uhr mit der Dauer der Reise wächst.

2. Jede Geschwindigkeitsänderung gibt eine dauernde Vergrösserung der Uhrperiode. — Auch unannehmbar; denn wenn es so wäre, müsste die Uhr schliesslich langsamer gehen.

3. Eine Vergrösserung der Geschwindigkeit relativ zu K von 0 zu v vergrössert die Uhrperiode dauernd, und eine Verminderung der Geschwindigkeit von v zu 0 vermindert die Uhrperiode in demselben Verhältnis. — Dadurch könnte die schliessliche Standänderung wohl erklärt werden; aber die Annahme lässt sich nicht mit der vollkommenen Gleichstellung von K und K' vereinbaren. Denn eine Vergrösserung der Geschwindigkeit relativ zu K von 0 zu v ist doch eine Verminderung von v zu 0 relativ zu K'. (Auch die Annahme, dass die Richtung für die Gangänderung entscheidend sei, wäre mit der Gleichstellung der Koordinatensysteme unvereinbar).

Der Mangel an Kausalität ist doch nicht gleichbedeutend mit einem Mangel an logischem Zusammenhang in der Theorie. Dem Anschein nach entsteht wohl ein Selbstwiderspruch dadurch, dass ein mit der Uhr U_2 folgender Beobachter während der gleichförmigen Bewegung stets finden wird, dass U_1 langsamer geht als U_2, und nichts desto weniger schliesslich — gleichgültig wie lange die Reise gedauert hat — erkennen wird, dass U_2 zurückgeblieben ist. Dieser scheinbare Selbstwiderspruch lässt sich doch durch Betrachtung der Verhältnisse in den Akzelerationsperioden auflösen (siehe z. B. »H. A. Lorentz: Das Relativitätsprinzip«, 1914, S. 47).

und dem Verhalten der Elektronen bei Bewegungen im Aether die Aethertheorie mit Einsteins Lehre in Uebereinstimmung zu bringen vermochte, konnte man meinen, dass die Sache in Ordnung sei.

Dem Einwand, dass die Lorentz'schen Annahmen einen willkürlichen, nur für den Fall erfundenen Ausweg bezeichnen, kann man kein grosses Gewicht zuschreiben. Denn bei unserem Suchen nach Gesetzen, durch welche wir von dem kausalen Zusammenhang Rechenschaft ablegen können, müssen wir uns überhaupt von dem Vorliegenden leiten lassen, und wenn wir dann durch einfache Annahmen dieses Ziel erreichen, fühlen wir uns befriedigt. Der wirkliche Mangel, welcher dem Lorentz'schen Ausweg anhaftet, liegt an einer ganz anderen Stelle, nämlich in ernsthaften Schwächen des angewandten Erklärungsgrundes — des Weltäthers — selbst. Die Einführung dieses Mediums als physikalische Ursache war überhaupt ein bedenklicher Schritt von dem natürlichen Wege ab, auf dem wir den Ursachen nachspüren, und die negativen Ergebnisse der vielen Versuche, dem Aether bestimmte stoffliche Eigenschaften zuzuschreiben, mussten ihn als reale physikalische Ursache allmählich mehr und mehr verdächtig machen. Die Realität der Körper, welche wir sonst als Ursachen einführen, erkennen wir daran, dass das Betragen anderer Körper in gesetzmässiger Weise von ihren Lagen und Bewegungen relativ zu gewissen Raumabschnitten, die dadurch von der Umgebung unterscheidbar werden, abhängig sind. Selbst für die Luft, deren Unscheinbarkeit und Homogenität oft mit denen des Aethers zusammengestellt werden, haben wir Realitätskriterien dieser Art. Schon bei den gröberen Beobachtungen finden wir verschiedene Stellen einer Luftmasse durch verschiedene Dichte u. a. m.

voneinander gekennzeichnet; und ein tieferes Eindringen in die physikalischen Vorgänge hat uns gelehrt, dass die Luft aus kleinen, individuellen Raumabschnitten, den Molekülen, besteht, die wir als Träger der physikalischen Ursachen aller von der Luft ausgeübten Wirkungen ansehen müssen. Aber was den Aether betrifft, so hatte man hier kein Anzeichen einer solchen Individualisierung gefunden; er figurierte in der Physik nur als ein unterschiedsloses Kontinuum. Man kann sagen, dass er eigentlich nur als ein Name eines Weltkoordinatensystems angewandt wurde, auf das man die Bewegungen des Lichtes und der Elektronen bezog, dass man aber keine reale, für die Beobachtung zugängliche Bezugspunkte für die Bewegungen gefunden hatte. Schon hierdurch musste Zweifel daran entstehen, ob in der Tat solch ein besonderes Medium existiere. Die negativen Resultate von dem Michelson-Versuch u. a. mussten diesen Zweifel erhärten. Und obgleich die früher erwähnte Lorentz-Hypothese wohl rein formal eine auf dem Weltäther basierte kausale Auffassung der speziellen Relativitätstheorie ermöglichte, kompromittierte sie in der Tat noch ernsthafter den Aether als *vera causa*, weil sie es prinzipiell unmöglich machte, Ruhe und gleichförmige Bewegung in dem Aether durch Beobachtungen voneinander zu unterscheiden. Einerseits hat sie zur Voraussetzung, dass eins von den gegeneinander bewegten Koordinatensystemen der speziellen Relativitätstheorie vor den anderen durch Ruhe im Aether bevorzugt wird, andererseits schliesst sie aber prinzipiell jedes beobachtbare Kennzeichen einer solchen Bevorzugung aus. Es liegt unzweifelhaft an der Erkenntnis dieser Schwäche der Fundamente der Aethertheorie, dass Lorentz sich über die Realität des Aethers sehr vorsichtig ausdrückt; er kann

nicht umhin, sagt er, den Aether als »*endowed with a certain degree of substantiality*« zu betrachten (»H. A. LORENTZ: The Theory of Electrons«, 1909). Und es ist verständlich, dass Einstein und viele mit ihm den Schritt weiter gingen, den Weltäther als ein blosses Phantasieprodukt anzusehen und es als Pseudophysik zu betrachten, wenn man, um den scheinbaren Mangel an Kausalität zu verdecken, dieses Medium als Ursache einführt.

Demungeachtet kann man nicht annehmen, dass Einstein die Anschauung hegt, dass die Relativitätstheorie den Ursachsbegriff über Bord wirft. Bei seiner Begründung der allgemeinen Relativitätstheorie hat er es als einen Mangel an Kausalität sowohl der speziellen Relativitätstheorie als der Newton'schen Mechanik hervorgehoben, dass man »ein reales Etwas« vermisst, worauf man das verschiedene Verhalten der Körper gegenüber den »galileischen Koordinatensystemen« und anderen Koordinatensystemen zurückführen kann (siehe »Ann. d. Physik« 49, 1916 S. 771 und »A. EINSTEIN: Ueber die spezielle und die allgemeine Relativitätstheorie« 2. Aufl. 1917 S. 49). Zur Beleuchtung dieses Mangels benutzt Einstein u. a. ein Beispiel mit zwei frei im Raume schwebenden, relativ zueinander rotierenden flüssigen Körpern, von denen der eine sphärisch, der andere ein Rotationsellipsoid ist. In näherem Anschluss an Newtons bekannte Betrachtungen über die Rotation von mit Wasser gefüllten Eimern wollen wir uns hier lieber zwei solche an dem Nordpole der Erde aufgehängte Eimer denken, den einen in Ruhe relativ zur Erde, den anderen relativ zu dieser um eine vertikale Achse mit dem Uhrzeiger einmal in einem Sternentage rotierend (also in »Rotationsruhe« relativ zu den Fixsternen). In dem zweiten von den Eimern wird die Wasseroberfläche ein Teil einer

Kugelfläche um das Erdzentrum sein, während sie in dem ersten eine gewisse Abflachung haben wird. Nach Newtons Mechanik wird die Abflachung in dem ersten Eimer dadurch erklärt, dass »die Inertie« der Wasserteilchen gewisse Zentrifugaldrücke erzeugt. Aber auf die Frage, warum die Inertie in der relativ zu den Fixsternen, aber nicht in der relativ zur Erde rotierenden Wassermasse diese Wirkung hat, hat man sich gewöhnlich mit der Antwort begnügt, dass die erste »im Raume« rotiert, während die zweite in Rotationsruhe in diesem ist. Hierdurch wird dann Bewegung im Raume oder »absolute Bewegung« als physikalische Ursache eingeführt. Einstein will aber keine absolute Bewegung als *vera causa* annehmen, und in Anschluss an E. Mach (siehe »E. Mach: Die Mechanik in ihrer Entwicklung« 7. Aufl. S. 227) hat er »ferne Massen (und ihre Relativbewegungen gegen die betrachteten Körper)« als »Träger von prinzipiell beobachtbaren Ursachen« zu dem verschiedenen Verhalten der Körper in galileischen und anderen Koordinatensystemen angegeben.

Im Folgenden wollen wir als die kausale Relativitätsforderung die Forderung bezeichnen, dass die Bewegungen oder Lagen, welche wir als bestimmend für die physikalischen Vorgänge annehmen, weder auf den leeren Raum noch auf ein hypothetisches Weltmedium ohne irgend welche für die Beobachtung zugängliche Bezugspunkte bezogen werden dürfen, sondern dass es Bewegungen oder Lagen relativ zu erkennbaren Körpern (d. i Raumabschnitten, die von der Umgebung unterscheidbar sind) sein müssen — oder, anders ausgedrückt, dass man nur solche Körper als Träger der physikalischen Ursachen ansehen darf.

In den Bezeichnungen »die kausale Relativitätsforderung« und »das spezielle Relativitätsprinzip« wird das Wort

»Relativität« augenscheinlich in wesentlich verschiedenem Sinne benutzt. Durch das spezielle Relativitätsprinzip wird ohne kausale Begründung Aequivalenz (mit Rücksicht auf die Ausdrücke für die physikalischen Gesetze) zwischen gewissen Koordinatensystemen postuliert, und man könnte es darum eigentlich treffender als ein Aequivalenzprinzip oder -postulat bezeichnen; wir wollen es Einsteins erstes Aequivalenzprinzip nennen. Einstein zieht aber aus den oben erwähnten kausalen Betrachtungen die Folgerung, dass »die Gesetze der Physik so beschaffen sein müssen, dass sie in Bezug auf beliebig bewegte Bezugssysteme gelten« (Ann. 49. S. 772). In diesem Satze begegnen die beiden verschiedenen Arten von Relativität, wie es scheint, einander, indem die Kausalitätsforderung zur Aufstellung von einem Aequivalenzsatze führt.

Man könnte hiernach erwarten, dass die Kausalität durch die Entwicklung der allgemeinen Relativitätstheorie wieder zu Ehren kommen würde, und dass dann auch die spezielle Relativitätstheorie, die ja eben nur ein spezieller Fall von der allgemeinen ist, in kausaler Beleuchtung hervortreten würde. Ich glaube auch, dass es in der Tat möglich ist, den realen Inhalt der allgemeinen Relativitätstheorie in einem solchen Lichte zu betrachten, dass die Kausalitätsforderung für diese — und somit auch für die spezielle Relativitätstheorie — als erfüllt erscheinen wird. Dass dies in den Einstein'schen Arbeiten nicht klar zu Tage zu kommen scheint, mag u. a. darauf beruhen, dass das erwähnte kausale Argument für Einstein nicht das Hauptargument war. Das Hervorziehen von »dem realen Etwas«, das die Basis einer kausalen Betrachtung bilden muss, scheint bei der Entwicklung der Relativitätstheorie für ihn nicht die Hauptsache gewesen zu sein; man kann

vielmehr sagen, dass im Gegenteil seine Theorie ihrem Wesen zufolge von der Betonung der Kausalität hinwegführt.

Wenn es nun scheint, dass das Einstein'sche Kovarianzprinzip (»Die allgemeinen Naturgesetze sind durch Gleichungen auszudrücken, die für alle Koordinatensysteme gelten« Ann. 49. S. 776) sich aus der kausalen Relativitätsforderung ableiten lässt, so hat es damit seine eigene Bewandtnis. Wenn es möglich wäre, für die physikalischen Gesetze solche auf ein gewisses Koordinatensystem bezogenen Ausdrücke aufzustellen, dass sie für alle möglichen gegenseitigen Lagen und Geschwindigkeiten aller in Betracht kommenden Körper gölten, so müssten allerdings der kausalen Relativitätsforderung zufolge diese Ausdrücke von der Wahl des Koordinatensystems unabhängig sein; denn sonst wären nicht nur reale, beobachtbare Dinge, sondern auch Koordinatensysteme für die physikalischen Vorgänge bestimmend. Es liegt aber in der kausalen Relativitätsforderung nichts, was uns sagt, dass dieses überhaupt möglich sein muss; von vorn herein scheint ein Versuch in dieser Richtung im Gegenteil ganz aussichtlos zu sein. Denn wenn auch das Weltgeschehen ein Wechselspiel zwischen individuellen Körpern von nur ganz wenigen (zwei?) Arten ist und der Einfluss von jedem einzelnen auf den Zustand in dem umgebenden Raume einfachen Regeln folgt, muss man doch erwarten, dass die willkürlichen und verschiedenen Relationen von allen einzelnen Körpern zu dem Koordinatensysteme diese Einfachheit ganz verschleiern werden. Die Verhältnisse werden überdies dadurch kompliziert, dass wir nicht die Einflüsse der einzelnen Körper als momentan sich fortpflanzende Fernwirkungen auffassen dürfen.

Dass es trotzdem Einstein gelungen ist, allgemein kovariante Ausdrücke für allgemeine physikalische Gesetze

aufzustellen, muss somit als ein höchst überraschendes Resultat bezeichnet werden. Von entscheidender Bedeutung für diese Leistung sind selbstverständlich die von Einstein angewandten, von jedem früheren Verfahren abweichenden Methoden gewesen. (Einführung von einem nicht-euklidischen Ausdruck für das Linienelement, Anwendung von vierdimensionalen Raum-Zeit-Koordinaten u. s. w.). Aber es kann durchaus nicht als eine notwendige Konsequenz des Kausalitätsprinzips oder überhaupt von erkenntnistheoretischen Betrachtungen angesehen werden, dass man ein solches Resultat erreichen musste, wenn man nur den rechten Weg einschlug; es war zugleich von einer Einfachheit besonderer Art in dem Weltgeschehen bedingt.

Indem Einstein von gewissen bekannten mechanischen und elektrodynamischen Aequivalenzverhältnissen und von der »Gleichheit der trägen und der schweren Masse« ausging, hat er eine solche tiefgehende Einfachheit in den physikalischen Erscheinungen aufgespürt und enthüllt, und in genauer Anknüpfung daran hat er zugleich eine neue fundamentale Methode für die analytische Behandlung der Erscheinungen entwickelt und dadurch neue grosse Möglichkeiten für unser Eindringen in deren Zusammenhang eröffnet. Aber der Umstand, dass das kausale in dem Geschehen in der Relativitätstheorie verschleiert wird, hat die erkenntnistheoretische Seite der Sache in einer für das volle Verständnis und die rechte Beurteilung der Theorie vielleicht nicht fördernden Weise in den Vordergrund gebracht. Durch viele Darstellungen der neuen Lehre bekommt man den Eindruck, dass sie einen entscheidenden erkenntnistheoretischen Bruch mit älteren Anschauungen bezeichnet, und während einige deswegen die Lehre abweisen, gibt

es andere, die eben darin ihre wahre Grösse sehen. Deshalb mag es nützlich sein, die Aufmerksamkeit besonders auf das Kausale zu richten, das unter den mathematischen Transformationen versteckt liegt, und dadurch zu erstreben, nicht den Bruch, sondern die Kontinuität gewahr zu werden.

In dieser Absicht wollen wir im Folgenden untersuchen, zu welchen Konsequenzen wir natürlich geführt werden, wenn wir die Erfüllung der kausalen Relativitätsforderung verlangen, im übrigen aber auf dem alten Grunde verharren. Wir wollen versuchen, nachzuweisen, dass man auf diesem Wege — also ohne die für Einsteins Theorie eigentümlichen Methoden und Annahmen einzuführen — was das Qualitative betrifft, durch einfache Betrachtungen ungezwungen zu mehreren Resultaten der Relativitätstheorie geführt wird, und das zwischen Einsteins Lehre und der älteren Physik kein im Wesen dieser Theorie notwendig begründeter Widerspruch zu bestehen scheint.

II.

Das Newton'sche Trägheitsgesetz hat nur dann einen bestimmten Sinn, wenn vorausgesetzt wird, dass die Bewegung auf gewisse Koordinatensysteme bezogen wird, die gegeneinander gleichförmige und geradlinige Translation haben. Denn ein Körper, der sich in einem bestimmten Koordinatensysteme gleichförmig und geradlinig bewegt, wird sich in einem anderen, das relativ zu dem ersten eine ungleichförmige Bewegung hat, ungleichförmig bewegen und im allgemeinen eine andere Bahn als eine gerade Linie beschreiben. Einstein nennt die Koordinaten-

systeme, in welchen (d. i. in bezug auf welche) das Trägheits- oder Inertialgesetz gilt, die galileischen; wir wollen sie als die Inertialsysteme bezeichnen. Auf diese soll man sowohl die elektrodynamischen als die mechanischen Gesetze beziehen. Dass das elektrodynamische Bezugssystem (»das Weltätherkoordinatensystem«), auf das die Elektronenbewegungen zu beziehen sind, ein Inertialsystem ist, lässt sich übrigens aus der Annahme des elektrischen Ursprunges der Trägheit ableiten.

Diese Bevorzugung von gewissen Koordinatensystemen lässt sich nur mit der kausalen Relativitätsforderung vereinbaren, wenn sie eine prinzipiell beobachtbare Realität vertreten — z. B. ein System X von Körpern, die nicht explicite in den Gesetzen vertreten sind, aber für das Geschehen in dem ins Auge gefassten System S mitbestimmend sind. Die für das System S geltenden Gesetze können allerdings dann keine allgemeinen sein; man muss annehmen, dass ihre Einfachheit durch eine Einfachheit der Relation von S zu X bedingt ist (und dass sie deswegen auch nur mit Annäherung gelten) — z. B. dass S sich innerhalb einer Raumgegend von kleinen Dimensionen relativ zu seinem Abstand von X befindet.

Es war bekannt, dass die Newton'schen Inertialsysteme relativ zur Erde rotierten, aber relativ zu einem sehr grossen äusseren Systeme — dem Fixsternensysteme — in Rotationsruhe waren; man meinte aber nur, dass dieses sowie die Inertialsysteme in Rotationsruhe im Raume seien (siehe S. 10). Wenn wir nun Ruhe und Bewegung im Raume als Erklärungsgrund abweisen, liegt die Annahme auf der Hand, dass die Sternenwelt — die sichtbaren Fixsterne und vielleicht noch grössere Sternensysteme — wirklich für die Inertialsysteme bestimmend sind. Ja, es kann

es behauptet werden, dass diese (von E. Mach aufgestellte) Annahme die einzig mögliche ist, wenn die kausale Relativitätsforderung erfüllt sein soll und wir bei unserm Suchen nach Ursachen nicht die Grundlage verlassen wollen, die unsere bisherigen Erfahrungen uns gegeben haben, denn wir kennen kein anderes, beobachtbares System, das wir als Erklärungsgrund benutzen können. Es gibt aber auch — ausser dem wichtigen Indizium, das von der Rotationsruhe der Inertialsysteme gegen das sichtbare Sternensystem geliefert wird — andere kräftige Gründe für eine solche Annahme.

Wir wissen, dass die Sterne Weltkörper sind, welche die bekannte Gravitationswirkung ausüben. Aber, wie Einstein hervorgehoben hat, ist »die Gleichheit von der trägen und der schweren Masse«[1] ein sehr kräftiges Argument für einen genauen Zusammenhang zwischen Gravitation und Trägheit und somit für die Annahme, dass die Sterne, indem sie Gravitationswirkung (im gewöhnlichen Sinne) ausüben, auch für die Trägheit Bedeutung haben.

Eine Betrachtung von mehr erkenntnistheoretischer Art hinsichtlich Stoff und Kraftfelder wird uns auch zu dem Gedanken führen, dass die Sternenwelt alle sowohl mechanischen als elektrodynamischen Vorgänge beeinflusst.

Die Tatsache, dass in dem Raume um einen Stern ein Gravitationsfeld existiert, wurde früher gewöhnlich in einer gar zu inhaltlosen und mathematischen Weise aufgefasst.

[1] Mann kann sagen, dass das Verhältnis, worauf Einstein mit diesem Ausdrucke hinzielt, nichts anderes ist als die Tatsache, dass alle Körper an demselben Orte immer dieselbe Gravitationsakzeleration erfahren. Dieses kann aber auch so ausgedrükt werden, dass die Gravitationskräfte, welche an einem gewissen Orte auf verschiedene Körper wirken — im Gegensatz zu Kräften anderer Art — immer der trägen Masse der Körper proportional sind.

Die meisten sahen hierin nur einen Ausdruck davon, dass ein Körper hier eine durch die Entfernung des Sternes und eine gewisse denselben charakterisierende Eigenschaft (seine Masse) bestimmte Beschleunigung gegen ihn hat, aber der Raum wurde als »leer« aufgefasst. Viele konnten sich doch nicht mit diesem Gedanken versöhnen; sie fanden es unbegreiflich, dass der Körper den Stern so trotz seiner Entfernung empfinden konnte; sie meinten deshalb, dass ausser dem Sterne und dem Körper ein unbekannter Stoff, ein Medium (der Aether), existieren müsste, welcher den sonst leeren Raum ausfüllte und die Fernwirkung zwischen ihnen vermittelte. Hierdurch könnten dann die sonst unfassbaren »Fernkräfte« erklärt werden. Ganz anders war die Auffassung von den »Berührungskräften«. Wenn ein ausgeworfener Ball von einer Wand angehalten wurde, war dieses verständlich, weil er in einen »stoffgefüllten Raum« eindrang (oder einzudringen versuchte); man fand hier kein besonderes Medium erforderlich, um die Sache verstehen zu können. Diese Vorstellung von unbegreiflichen Fernkräften, die ein besonderes Medium zu ihrer Erklärung erfordern, und begreiflichen, oder wenigstens begreiflicheren Berührungskräften, ist aber nicht nur erkenntnistheoretisch unhaltbar, sondern auch ohne Wurzeln in der gegenwärtigen Physik — was doch nicht verhindert, dass sie noch in grossem Umfang die Auffassung von physikalischen Erscheinungen beeinflusst.

Dasjenige, worin »der Ball« eindringen will, ist in der Tat nichts anderes als die eigentümlichen Kraftfelder ganz dicht an den Atomen und in ihrem Innern (d. i. in Entfernungen von ihren Kernen und Elektronen, die von einer mit den gegenseitigen Entfernungen zwischen diesen vergleichbaren Grössenordnung sind), und innerhalb dieser

Kraftfelder variiert die Kraft kontinuierlich wie in den Gravitationsfeldern der Weltkörper. Was wir einen Körper oder einen stoffgefüllten Raumabschnitt nennen, ist nur ein Raumabschnitt, der mit einer Sammlung von konzentrierten, von Ort zu Ort stark variierenden Kraftfeldern eigentümlicher Art gefüllt ist; aber der Raum zwischen den Stoffabschnitten der Weltkörper ist ebenso wohl von Kraftfeldern ausgefüllt, wenn diese auch von anderer Art sind. Der Unterschied kann physikalisch dadurch charakterisiert werden, dass wir in den Stoffabschnitten mit elektrischen Feldern zu tun haben, indem Atomkerne und Elektronen hier ihren Einfluss je für sich geltend machen können (siehe die vorhergehende Parenthese), während wir ausserhalb der Stoffabschnitte »Neutralfelder« haben, worin ihr Einfluss verschmolzen ist. Wenn man nicht alle Kraftfelder »Stoff« nennen will — was rationell, aber unpraktisch sein würde — aber doch diesem Begriff eine rationelle Abgrenzung geben will, gibt es somit keinen andern Weg als den Einstein'schen: nicht nur Stoff im gewöhnlichen Sinne, sondern jedes elektromagnetische Feld Stoff zu nennen (Annalen 49. S. 803). Aber der in erkenntnistheoretischer Hinsicht tiefgehende Unterschied zwischen gefülltem und leerem Raum ist der Unterschied zwischen Raum mit Kraftfeldern und Raum ohne Kraftfelder. Eine Unterscheidung, wonach ein Raum als stoffgefüllt betrachtet wird, wenn die Bewegung eines Körpers Widerstand erfährt, aber als leer, wenn dieses nicht der Fall ist, lässt sich nicht rationell aufrechterhalten. In einem von zahlreichen Weltkörpern erzeugten Neutralfeld, wird der Gravitationseinfluss von einigen unter diesen auf die Geschwindigkeit eines bewegten Körpers eine vergrössernde Wirkung haben, während andere von ihnen engegengesetzt

wirken. Das Resultat kann entweder eine Vergrösserung oder eine Verminderung sein. In letzterem Falle kann man sagen, dass die Bewegung einen Widerstand erfährt — ebensowohl als wenn die Bewegung in »Stoffgebieten« gehemmt wird. — In der Bedeutung »frei von Kraftfeldern« aufgefasst, wird »leer« nur ein Grenzbegriff; der leere Raum existiert nicht in der Welt, die wir kennen.

Man kann den Tatbestand so ausdrücken, dass ein Weltkörper ein Individuum ist, das aus einem Stoffabschnitte und einem Neutralfelde besteht. Dieses letztere, das sich ins Unendliche erstreckt, ist ein ebenso integrierender Bestandteil von dem Individuum wie der Stoffabschnitt, und es ist darum berechtigt, zu sagen, dass ein Weltkörper (und überhaupt jeder Körper) unendlich gross ist oder die ganze Welt erfüllt, aber in solcher Weise, dass das Feld immer schwächer wird, d. h. sein Einfluss auf das physikalische Geschehen immer geringer wird, je grösser die Entfernung von dem Stoffabschnitte ist — oder, mit andern Worten, dass die Intensität des Feldes mit der Entfernung von dem Stoffabschnitte abnimmt. Ein Weltkörper ist ferner ein zusammengesetztes Individuum, und sowohl sein Stoffabschnitt als sein Neutralfeld sind aus den elektromagnetischen Feldern und den Neutralfeldern zusammengesetzt, aus denen die einzelnen Atome bestehen. Wenn man, obschon die Individuen in dem Sinne, worin wir das Wort hier nehmen, unendlich gross sind, doch von ihren relativen Lagen und Bewegungen sprechen kann, liegt das daran, dass Veränderungen in und Bewegungen von dem kleinen Stoffabschnitt für den Einfluss des ganzen Individuums auf andere Individuen bestimmend sind.

Die Erkenntnis, dass der Raum zwischen den Welt-

körpern von etwas ebenso Realem wie dem, was wir Stoff nennen, erfüllt ist — nämlich dem von den Neutralfeldern von nahen und fernen Weltkörpern erzeugten zusammengesetzten Neutralfeld — kann nicht umhin, unsere Auffassung von den physikalischen Vorgängen im Raume zu beeinflussen. Das Licht z. B. kann als eine periodische, vorschreitende Aenderung in diesem Raumzustande aufgefasst werden, und wir haben darum ebenso wenig Grund, die Existenz eines besonderen, für diesen Zweck erfundenen Mediums anzunehmen, als die Fernkräfte durch eine solche Annahme erklären zu wollen. Die Hypothese von einem besonderen Weltmedium erscheint denn von vorn herein als überflüssig, weil wir dadurch im Verständnis von dem Weltgeschehen um keinen Schritt weiter kommen, als wenn wir die erkennbaren Körper als Träger von allen Ursachen und die durch sie bedingten Raumzustände als für alles Geschehen bestimmend ansehen. Diese Annahme, nach welcher sowohl die elektrodynamischen als die mechanischen Vorgänge, sowohl die Trägheit als die Gravitation, in einer gewissen Raumgegend von diesem realen Raumzustand in der betreffenden Gegend abhängig sind, erscheint dagegen als eine natürliche Hypothese.

Einige könnten vielleicht den Eindruck bekommen, dass man sich durch die hier entwickelte Auffassung in gewissem Masse der alten Vorstellung von unvermittelten Fernwirkungen anschliesst. Das trifft jedoch nicht zu. Teils wird der Begriff »Fernwirkung« eigentlich ausser Rede gestellt dadurch, dass die Individuen je für sich als unendlich gross oder als einen sich über den ganzen Raum erstreckenden Raumzustand darstellend angesehen werden, teils werden weitere Spekulationen darüber, wie das jedem Individuum gehörende Feld zu Stande kommt, durch die

angestellten Betrachtungen nicht ausgeschlossen. Was hier abgelehnt wird, ist ja nur der alte Streit, ob man direkte Fernwirkungen oder Wirkungen durch ein Medium vorziehen will. Die Fernwirkungsfrage hat aber andere Seiten. In realer Hinsicht ist es eine Frage, ob Aenderungen in oder Bewegungen von dem Stoffabschnitte eines Individuums momentan ihren Einfluss über die ganze Welt geltend machen, oder ob die Wirkung sich mit endlicher Geschwindigkeit fortpflanzt. Die letztere mit den modernen Anschauungen übereinstimmende Annahme steht keineswegs mit der entwickelten Auffassung in Widerspruch. In formaler (rechnerischer) Hinsicht ist es eine Frage, ob man in die Formeln der Physik die einzelnen Individuen — durch die Koordinaten ihrer Stoffabschnitte und gewisse Eigenschaftsgrössen (Masse, elektrische Ladungen u. s. w.) charakterisiert — einführen will, oder ob man (wie in Differentialtheorien) Grössen, wodurch der von sämtlichen implizierten Individuen bestimmte Zustand in einem gewissen Punkte charakterisiert wird, einführen will.

III.

Es muss angenommen werden, dass die Beschleunigung, welche die Körper in einem Gravitationsfelde (im alten Sinne)[1] in einer gewissen Richtung erfahren, von einer Ungleichartigkeit der Intensität des Neutralfeldes bedingt ist. Eine solche Ungleichartigkeit hat man in dem Neutralfeld eines jeden Weltkörpers, weil die Intensität des Feldes mit der Entfernung von dem Stoffabschnitte des

[1] Um Zweideutigkeit zu vermeiden, wollen wir das Wort »Gravitationsfeld« nur in dieser Bedeutung benutzen und für den ganzen Zustand die Bezeichnung »Neutralfeld« anwenden.

Weltkörpers abnimmt, und wenn ein Körper an einem gewissen Orte zwischen zwei Weltkörpern keine Beschleunigung gegen den einen oder den anderen von ihnen erfährt, findet dieses Verhältnis seine hinlängliche Erklärung darin, dass die Intensität hier einen Minimalwert hat und deshalb nicht in der einen oder der anderen Richtung zunimmt. Es würde aber unberechtigt sein, daraus zu schliessen, dass die beiden Neutralfelder dann einander aufgehoben haben, so dass der Zustand des Raumes hier der nämliche ist, als wenn sie nicht existierten. Im Gegenteil muss angenommen werden, dass der Einfluss von beiden Weltkörpern zusammen auf Vorgänge, die nicht wie die Beschleunigung von Differenzen in der Intensität bedingt sind, grösser ist als der Einfluss von jedem einzelnen der Weltkörper.

Wenn die einzelnen Weltkörper, die ein zusammengesetztes Neutralfeld bilden, sich relativ zueinander bewegen, wird das Feld in einem Aenderungszustand sein; denn die individuellen Neutralfelder verschieben sich ineinander. Wir wollen aber annehmen, dass man sich immer für einen gewissen Ort und einen gewissen Zeitpunkt ein solches Feld mit Rücksicht auf seinen Einfluss auf die physikalischen Vorgänge durch ein Neutralfeld ohne innere Verschiebungen ersetzt denken kann. Hierin liegt, dass man sich immer ein feldbildendes System, das aus einer Anzahl von relativ zueinander bewegten Weltkörpern besteht, für eine genügend kleine Raumgegend und eine genügend kleine Zeit durch ein System von feldbildenden Körpern in gegenseitiger Ruhe oder mit anderen Worten durch einen einzelnen festen Körper ersetzt denken kann. Wir wollen uns hier damit begnügen, vorauszusetzen, dass es prinzipiell möglich ist, ein solches »Ersatzsystem«

oder einen solchen »Ersatzkörper« durch Anwendung von bestimmten Gesetzen zu bestimmen; was das für Gesetze sind, ist eine Frage, die ausserhalb des Zwecks dieser Abhandlung liegt. Die genannte Voraussetzung ist allerdings eine Hypothese; aber die entgegengesetzte Annahme würde eigentümliche Konsequenzen mit sich führen. Die gegenseitigen Bewegungen der feldbildenden Sterne würden dann nicht nur zeitliche Aenderungen in dem Zustand an verschiedenen Orten verursachen, sondern auch besondere Raumeigenschaften erzeugen, die ganz wegfallen würden, wenn die feldbildenden Sterne in gegenseitiger Ruhe wären. Wir haben aber keinen Erfahrungsgrund, der uns dazu führt, eine solche Komplikation anzunehmen.

Wenn wir die Verhältnisse in unseren Raumgegenden betrachten, so haben wir hier an dem Sonnensysteme ein materielles System S von sehr geringen Dimensionen relativ zu seinen Entfernungen selbst von den nächsten bedeutenden äusseren Weltkörpern, die das felderzeugende Fixsternensystem X bilden. Das zusammengesetzte Neutralfeld von diesem System bildet, so zu sagen, eine Atmosphäre, worin das Sonnensystem sich befindet, und alle physikalischen Vorgänge in unseren Raumgegenden werden teils durch die Wechselwirkungen innerhalb des Systems, teils durch diese Atmosphäre bestimmt. Da die relativen Bewegungen, welche die Fixsterne selbst in langen Zeiträumen ausführen, relativ zu den enormen Dimensionen des Fixsternensystems sehr klein sind, kann man sich für die ganze Raumgegend des Sonnensystems das Neutralfeld der Sterne durch einen festen »Sternenkörper« — eine mächtige feste »Welthülle« von Stoff, die das Sonnensystem in grossem Abstande umgibt, erzeugt denken. Ein Koordinatenssystem, das im Neutralfelde, d. i. relativ

zu dieser Welthülle, ruht, wollen wir als das »Ruhekoordinatensystem«[1] (mit Rücksicht auf die äussere felderzeugende Welt) für die betreffende Raumgegend bezeichnen. Obgleich es den Einfluss von der ganzen Sternenwelt vertritt, ist es doch kein Weltkoordinatensystem in universalem Sinne. Es ist wohl das Ruhekoordinatensystem für eine Raumgegend von sehr grosser Ausdehnung, aber ensprechende Ruhekoordinatensysteme für andere, fernere Weltgegenden können sowohl in Rotation als in Translation (gleichförmig oder ungleichförmig) relativ zu dem unsrigen sein.

Es lässt sich denken, dass das von der Sternenwelt gebildete Neutralfeld in unseren Raumgegenden überall die nämliche Feldintensität hätte, oder dass wir uns hier in — wie wir sagen möchten — »einem homogenen Neutralfeld« befänden. Wenn ein sich selbst überlassener Körper in einem solchen Feld in Ruhe wäre, müssen wir annehmen, dass er in Ruhe verbleiben würde, da wir keine Ursache zu einer Aenderung dieses Zustandes sehen können. Das Ruhekoordinatensystem muss dann ein Inertialsystem für die betreffende Gegend sein, vorausgesetzt, dass ein Inertialsystem überhaupt existiert. Dass dieses der Fall ist, kann doch nicht als eine Selbstfolge oder eine Folge der kausalen Relativitätsforderung betrachtet werden. Denn wenn ein Körper, der sich in einem homogenen Felde bewegte, immer eine von der Geschwindigkeit abhängige Verzögerung erführe, könnten wir für dieses Verhalten seine Bewegung relativ zu erkennbaren Körpern (den felderzeugenden Ster-

[1] Da die Lage des Anfangspunktes und die Richtung der Achsen in den Koordinatensystemen, von welchen wir im Folgenden reden, für uns von keinem Belang sind, wollen wir alle relativ zueinander ruhenden Koordinatensysteme als ein einzelnes besprechen.

nen) als Ursache oder als bestimmenden Faktor in einer kausalen Beschreibung angeben. — Wenn das Ruhekoordinatensystem ein Inertialsystem ist, muss auch ein jedes relativ dazu gleichförmig und geradlinig bewegtes Koordinatensystem ein Inertialsystem sein. Wir können das im Felde ruhende als »das primäre Inertialsystem«, die anderen als sekundäre Inertialsysteme bezeichnen. Das Feld selbst können wir auch ein »gravitationsfreies Feld« oder ein »Inertialfeld« benennen.

Es ist jedoch eine ganz willkürliche Annahme, dass das Feld der Sternenwelt diese Eigenschaft in unserer Raumgegend besitzt. Es muss angenommen werden, dass die Verteilung der Sterne eine solche ist, dass die Intensität des Feldes in irgend einer Richtung zunimmt, und Körper, die sich selbst überlassen sind, werden dann in dieser Richtung Beschleunigung erfahren. Wegen der grossen Entfernungen der Sterne darf man aber davon ausgehen, dass die Beschleunigung an Grösse und Richtung über eine sehr grosse Raumgegend sehr annäherungsweise dieselbe ist; man kann mit anderen Worten das Feld als »ein homogenes Gravitationsfeld« ansehen. Unter der Voraussetzung, dass Körper jeder Art dieselbe Beschleunigung erfahren, wird dann jeder Körper in einem Koordinatensystem, das in dem Ruhekoordinatensysteme diese für das Feld charakteristische Beschleunigung hat, sich gleichförmig und geradlinig bewegen. Ein solches Koordinatensystem ist folglich ein Inertialsystem. Das heisst, wir können den Schluss ziehen, dass das Trägheitsgezetz für ein solches Koordinatensystem gilt. Dass Inertialsysteme verschiedener Art nicht nur diese Eigenschaft gemeinsam haben, sondern dass alle physikalischen Vorgänge in allen Inertialsystemen genau gleich verlaufen, ist keine notwen-

dige Folgerung: es ist dies eine Annahme, die in Einsteins Aequivalenzprinzipien ausgedrückt ist. Der ganze Inhalt von sowohl dem speziellen Relativitätsprinzip als »der Aequivalenzhypothese« (Einsteins erstem und zweitem Aequivalenzprinzip) lässt sich eben durch den kurzen Satz: »Alle Inertialsysteme sind äquivalent« ausdrücken. — Wir wollen ein Koordinatensystem der genannten Art »ein feldakzeleriertes Inertialsystem« nennen. Auch unter diesen Systemen wird in einem gewissen Augenblick eins, das primäre System, dadurch bevorzugt, dass es die Geschwindigkeit Null im Felde hat; aber es werden wegen der Beschleunigung unablässig neue Systeme sein, die diese Bevorzugung bekommen.

Wir haben hier nur das von der äusseren Sternenwelt gebildete Neutralfeld und das Verhalten von isolierten Körpern in diesem Felde betrachtet. Den Einfluss von den Weltkörpern des Sonnensystems und andere Einwirkungen müsste man dann besonders in Betracht nehmen. Wir können aber wohl in das felderzeugende System die näheren Weltkörper — die Sonne, die Erde und die Planeten — mit aufnehmen. Wegen der grossen Nähe von diesen relativ zu der Sternenwelt in verschiedener Weise bewegten Weltkörpern könnte man vermuten, dass sie bei der Bestimmung von dem festen Ersatzsysteme, das für eine gewisse Raumgegend das felderzeugende System repräsentieren soll, von grossem Belang sein würden, und dass man darum ein Ruhekoordinatensystem nur für eine kleine Raumgegend und eine kleine Zeit festlegen könnte. Wenn wir nur mit der »Fixsternen-Welthülle« und der relativ zu dieser rotierenden Erde zu tun hätten, wäre es augenscheinlich nicht mit der kausalen Relativitätsforderung vereinbar, dass das Ruhekoordinatensystem ganz mit

der Sternenhülle folgte. Wenn diese nicht existierte, würde das ganze Neutralfeld relativ zur Erde ruhen, und wenn sowohl die Sternenhülle als die Erde zu dem Neutralfelde beitragen, muss das Ruhekoordinatensystem um so mehr in der Erdrotation mitfolgen, je näher wir der Erde sind. Aus der Tatsache, dass die irdischen durch Pendelversuche u. dgl. bestimmten Inertialsysteme, soweit sich feststellen lässt, in Rotationsruhe relativ zu dem Fixsternensysteme sind, muss man dann die Folgerung ziehen, dass der Beitrag der Erde zu dem Neutralfelde, selbst in ihrer unmittelbaren Nähe, im Vergleich mit dem Beitrage der Sternenwelt, nur sehr gering ist. Besonders genau ist die erwähnte Bestimmung doch nicht (die Fehlergrenze ist wohl von der Grössenordnung 1%); aber auch andere Gründe sprechen dafür, dass die Intensität von dem Neutralfelde der äusseren Sternenwelt im Vergleich mit den Beiträgen von den Weltkörpern des Sonnensystems überwältigend gross ist, und dass folglich das Mitaufnehmen von diesen in das felderzeugende System das Ruhekoordinatensystem für die verschiedenen Gegenden nur in sehr geringem Grade zu ändern vermag. Etwas anderes ist es, dass die kleinen lokalen Variationen, die sie in der überaus grossen Feldintensität erzeugen, Beschleunigungen in verschiedenen Richtungen veranlassen, wodurch wir verschiedene in dem Ruhesystem feldakzelerierte, lokale Inertialsysteme bekommen — z. B. die kleinen irdischen Inertialsysteme, die von frei fallenden Körpern repräsentiert werden. (Ein freifallender Elevator mit Inhalt ist ein materielles System S, für welches das Trägheitsgesetz in bezug auf ein relativ zu dem Elevator ruhendes Koordinatensystem gilt.)

Bisher haben wir nur die Konsequenzen betrachtet,

welche die kausale Relativitätsforderung mit sich führt mit Rücksicht auf die Auffassung von den physikalisch bevorzugten Koordinatensystemen, auf welche wir die Bewegungen beziehen müssen, um einfache Ausdrücke für die Bewegungsgesetze zu bekommen. Jetzt wollen wir auch einige Konsequenzen anderer Art betrachten, wobei wir der Einfachheit halber vorläufig nur an die Verhältnisse in homogenen Neutralfeldern denken.

Wenn wir das von den Weltkörpern erzeugte Neutralfeld an die Stelle des frühern Weltäthers setzen, können wir die Lorentz'sche Elektrodynamik für eine Gegend mit homogenem Neutralfeld anwenden. Elektronen, die im Felde (d. i. in dem Ruhekoordinatensystem) ruhen, werden nur ein elektrostatisches Feld bilden. Durch ihre Bewegungen in dem Neutralfelde werden magnetische Felder erzeugt, und durch die dazu erforderliche Energie wird die Trägheit in der bekannten Weise erklärt. Aber wenn wir annehmen, dass es die Bewegung im Neutralfelde ist, die für diese Wirkungen bestimmend ist, so muss man erwarten, dass letztere von der Intensität des Feldes abhängig sind und mit dieser abnehmen. In einem leeren (feldfreien) Raum kann die Bewegung kein magnetisches Feld erzeugen: die Masse eines Körpers sollte dann gleich 0 sein und somit seine Bewegung eine unbestimmte werden. Diese Konsequenz geht auch direkt aus der kausalen Relativitätsforderung hervor, und es kann nicht wundern, dass man zu absurden Resultaten kommt, wenn man sich einen Leerheitszustand vorstellt, welcher nur ein Grenzbegriff, kein realer, physikalischer Zustand ist.

Wenn sowohl die Kräfte zwischen den Körperteilen wie die Massen von elektrischer Natur sind, muss man erwarten, dass sowohl die Dimensionen fester Körper als

die Periodenlängen von periodischen Systemen mit der Intensität des Feldes variieren. Die Kontraktionen in der Bewegungsrichtung, welche Lorentz aus der Annahme von der elektrischen Natur der innern Kräfte hergeleitet hat, müssen ebenso mit der Feldintensität variieren; wenn diese sich 0 nähert, müssen die Verkürzungen sich auch dieser Grenze nähern; der Zusammenziehungsfaktor muss sich also 1 nähern.

Was das Licht betrifft, so können wir dieses nicht gut als eine periodische, vorschreitende Zustandsänderung in dem Neutralfelde ansehen, ohne dann auch anzunehmen, dass die Lichtbewegung in mehreren Weisen von dem Zustand des Feldes abhängt. Man muss es von vorn herein als im höchsten Grade wahrscheinlich ansehen, dass das Licht sich nur in einem homogenen Neutralfelde gleichförmig und geradlinig bewegt, dass es aber in Neutralfeldern mit variierender Intensität variierende Geschwindigkeit und Richtung hat. Ob die Geschwindigkeit abnehmen oder zunehmen wird, wenn die Intensität abnimmt, darüber lässt sich kaum unmittelbar irgend etwas schliessen; wir können aber das, was früher von dem Lorentz'schen Zusammenziehungsfaktor gesagt wurde, benutzen. Dieser ist, wie bekannt, $\sqrt{1-\left(\frac{v}{c}\right)^2}$, wo v und c beziehungsweise die Geschwindigkeit des Körpers und des Lichtes sind; wenn dieser Faktor sich 1 für jeden Wert von v nähern soll, muss $\frac{v}{c}$ sich 0, c sich also ∞ nähern. Die Lichtgeschwindigkeit im Felde sollte also zunehmen und sich ∞ nähern, wenn der Zustand des Raumes sich dem Leerheitszustand nähert. Hieraus folgt, dass ein Weltkörper, der eine lokale Vergrösserung von der Intensität des Feldes

verursacht, bewirken wird, dass vorbeipassierendes Licht eine Krümmung mit der konvexen Seite nach aussen erfährt. Wenn man diese — von Einstein aus der Aequivalenzhypothese hergeleitete — Krümmung bisher nicht wahrgenommen hat, so ist dies wieder ein Zeugnis davon, dass der Beitrag, welchen selbst ein Weltkörper wie die Sonne zum Neutralfelde gibt, gegen den der Sternenwelt sehr gering werden muss, wenn die ganze Betrachtung mit den Beobachtungen stimmen soll.

Dass auch die Bahnen, welche die Weltkörper des Sonnensystems beschreiben, von den Variationen des Feldes einigermassen beeinflusst werden müssen, folgt schon daraus, dass die Ruhe- und Inertialsysteme eine gewisse Störung erfahren. Kleine Abweichungen von den Newton'schen Gesetzen müssen also erwartet werden.

Im Vorhergehenden wurde vorausgesetzt, dass die Bildung der elektromagnetischen Felder in einem homogenen Neutralfelde stattfinde. In der Tat haben wir jedoch in den Gravitationsfeldern die elektrodynamischen Gesetze kennen gelernt, und man kann deshalb nicht annehmen, dass die Homogenität des Feldes entscheidend ist. Wenn wir aber fortwährend voraussetzen, dass es die Geschwindigkeit der Elektronen im Neutralfelde ist, die für die Magnetfelder bestimmend ist, und wir davon ausgehen, dass das Gravitationsfeld bei einem freien Fall die zu ihrer Erzeugung erforderliche Energie leistet, sind wir dadurch im wesentlichen zu dem früheren Fall zurückgeführt worden, da eine allen Teilen eines materiellen Systems gemeinsame Geschwindigkeit der Lorentz-Hypothese zufolge nicht in den Vorgängen innerhalb des Systems hervortreten soll. In einem Koordinatensysteme, das die nämliche Beschleunigung wie die frei fallenden Körper hat (ein

feldakzeleriertes Inertialsystem) werden dann die physikalischen Vorgänge trotz der zunehmenden Geschwindigkeit im Neutralfelde annäherungsweise in gleicher Weise wie in einem Inertialsystem in dem homogenen Feld verlaufen, wenn die allmähliche Vergrösserung der Feldintensität so klein ist, dass die davon herrührende Komplikation die Erscheinungen nicht merkbar beeinflussen kann. Dass diese Gleichheit mehr als eine Annäherung ist (wie aus Einsteins Theorie folgt), kann doch weder aus der kausalen Relativitätsforderung noch aus der Lorentz-Hypothese gefolgert werden.

Als wir das besondere Aethermedium und »den absoluten Raum« als physikalische Ursachen aufgaben und statt derselben das reale Feld, das im Raume zwischen den Weltkörpern existiert, einführten, wurden wir durch eine Reihe von Betrachtungen — welche sich uns fast mit Notwendigkeit aufdrängen, wenn wir uns auf den Standpunkt der kausalen Relativitätsforderung stellen — zur Annahme von Variationen (in Masstäben, Uhren, Lichtgeschwindigkeit, Massen u. a. m.) von derselben Art geführt wie die Variationen, die Einstein von seinen Aequivalenzprinzipien (oder wenigstens hauptsächlich von diesen) abgeleitet hat. An den Aequivalenzprinzipien hatte Einstein aber Leitregeln zum Auffinden vom quantitativen Resultaten in betreff der Variationen. Solche Resultate können selbstverständlich nicht von kausalen Betrachtungen allein geleistet werden. Damit soll doch nicht gesagt sein, dass die gewöhnlichen induktiv-deduktiven Wege der Physik auf kausaler Grundlage in diesen Gebieten überhaupt nicht fahrbar sind.

Die Annahme, dass die Felder der Weltkörper nicht nur für die Gravitation im alten Sinne, sondern für andere

Vorgänge im Raume bestimmend sind, kann selbstverständlich nur als eine Hypothese betrachtet werden, so lange nicht durch Beobachtungen gezeigt worden ist, dass Variationen in diesen Vorgängen wirklich der stellenweisen und ungleichen Verteilung der Weltkörper im Raum entsprechen. Es muss aber hervorgehoben werden, dass diese »Neutralfeldhypothese« einen allgemeinern Charakter hat als Einsteins Relativitätstheorie oder Aequivalenzhypothese. Eben weil diese ganz bestimmte quantitative Relationen mit sich führt, würde sie durch davon abweichende Ergebnisse hinfällig werden. Die mehr allgemeine Hypothese könnte mehr vertragen. Namentlich könnten rein negative Resultate des Suchens nach solchen Variationen nicht als Gegenbeweis dienen. Die Lichtkrümmungen und Spektralverschiebungen z. B., welche die Sonne hervorbringen sollte, könnten vielleicht so klein sein, dass es ganz unmöglich wäre, sie mit den feinsten Beobachtungsmitteln der Gegenwart nachzuweisen. Von vorn herein können wir nämlich nicht wissen, ob die Intensität des von der gesamten Sternenwelt erzeugten Neutralfeldes nicht so überwältigend ist, dass der Sondereinfluss der Erde, der Sonne oder anderer Weltkörper, für welche Untersuchungen in dieser Richtung sich überhaupt anstellen lassen, ein ganz unmerklicher ist. Selbst wenn es so wäre, dass die Beobachtung uns keine Stütze für die Hypothese gäbe, wäre es doch vorzuziehen, die realen Körper im Weltraume als die erforderlichen Ursachen anzusehen, anstatt ein besonderes, unbeobachtbares Weltmedium dazu zu erfinden. Die Ergebnisse und Verheissungen der Einstein'schen Theorie machten es aber wahrscheinlich, dass die Neutralfeldhypothese in dem Realen eine sehr solide Verankerung finden wird.

Was speziell die Perioden der von den Atomen ausgesandten Lichtschwingungen betrifft, so muss erkannt werden, dass man nicht aus der Neutralfeldhypothese ohne weiteres schliessen kann, dass sie mit der Feldintensität variieren müssen. Die Möglichkeit lässt sich nicht von vorn herein abweisen, dass die Quantenphänomene, welche man nicht aus der klassischen Elektrodynamik ableiten kann, in dieser Hinsicht eine Ausnahmestellung einnehmen. Wenn man Lichtkrümmungen und Abweichungen von den Newton'schen Gesetzen findet, welche mit der Einstein'schen Relativitätstheorie übereinstimmen, wird man deshalb wohl solche Ergebnisse als Bestätigungen der Neutralfeldhypothese ansehen können, wenn auch durch Beobachtungen bewiesen würde, dass die Einstein'schen Spectralverschiebungen nicht stattfinden. Entscheidende negative Ergebnisse in dieser Hinsicht würden dagegen für das grundwesentliche der Einstein'schen Relativitätstheorie ein ernsthafter Schlag sein. Es ist Grund dazu vorhanden, diese Sache zu berühren, da gewisse Beobachtungen über das Sonnenspektrum, wie es scheint, mit Einsteins Theorie von den Spektralverschiebungen unvereinbar sind (C. E. St. John; »Astrophysical Journal« Vol. 46 S. 249). A. S. Eddington sagt hiervon, dass es nutzlos sein würde, den Ernst dieses scheinbaren Bruchs von Einsteins Theorie zu leugnen, dass wir aber vielleicht wegen der Schwierigkeiten der Untersuchungen das Urteil aufschieben dürfen (Eddington: »Report on the Relativity Theory of Gravitation«; The Physical Society, London; S. 57). Diesen Standpunkt einzunehmen, ist unzweifelhaft vorläufig notwendig.

IV.

Der Umstand, dass es durch die im Vorhergehenden angestellten kausalen Betrachtungen — die übrigens in mehreren Punkten mit gewissen Einstein'schen Betrachtungen zusammenfallen — möglich gemacht wurde, eine Strecke in der nämlichen Richtung vorwärts zu kommen, in der Einstein auf seinen eigenen Wegen so weit gelangt ist, muss die Meinung, dass Einsteins Relativitätstheorie in keinem realen Gegensatz zum Kausalitätsprinzip steht, kräftig erhärten. Wir müssen aber ein wenig näher untersuchen, woran es liegt, dass es aussehen kann, als ob der Gegensatz zugegen wäre.

Von entscheidender Bedeutung in dieser Hinsicht ist es, dass die Relativitätstheorie in ihrer Eigenschaft als Aequivalenztheorie vor allen Dingen darauf hinzielt, verschiedene Koordinatensysteme als äquivalent oder für die Beschreibung der physikalischen Vorgänge gleichberechtigt aufzustellen. Einsteins erstes Aequivalenzprinzip (das spezielle Relativitätsprinzip) stellt verschiedene relativ zueinander gleichförmig bewegte Inertialsysteme als äquivalent, und sein zweites Aequivalenzprinzip (»die Aequivalenzhypothese«) stellt Koordinatensysteme, die in einem gravitationsfreien Felde akzeleriert sind, als äquivalent mit Koordinatensystemen auf, die in einem homogenen Gravitationsfelde ruhen — oder umgekehrt Koordinatensysteme, die in einem Gravitationsfelde feldakzeleriert sind, als äquivalent mit Koordinatensystemen, auf, die in einem gravitationsfreien Felde gleichförmig bewegt sind. Daher steht es einigermassen in Widerspruch mit dem Zweck der Theorie, oder es liegt wenigstens ausserhalb ihres Planes, äussere physikalische Verhältnisse, welche die Koordinatensysteme ungleich stel-

len können, in volle Beleuchtung zu bringen, anstatt auf die innere phänomenologische Gleichheit Gewicht zu legen. Da es mehr eine formale als eine reale Frage ist, können die Ausdrucksformen grosse Bedeutung erhalten, und es ist unzweifelhaft, dass viele Missverständnisse, viele Desorientierungen in betreff der Relativitätstheorie nur durch die Darstellungsweise verursacht worden sind. Wenn man z. B. den Ausdruck anwendet, dass ein Gravitationsfeld durch eine Koordinatenwahl »erzeugt« oder »zum Verschwinden gebracht« wird, kann dies als eine Art von Zauberkunst erscheinen, durch die man mittels eines mathematischen Kunstgriffs eine physikalische Verwandlung ausführt, wenn es auch möglich ist, darin nur eine zulässige Gedankenoperation zu sehen.

Wenn es der Zweck ist, die kausale Seite der Sache zu beleuchten, muss man aber im Gegenteil die physikalischen Verhältnisse hervorziehen, welche gewisse Koordinatensysteme von anderen unterscheiden, die diesen in der Relativitätstheorie äquivalent gesetzt werden. Zum Teil eben aus diesem Grunde wurden im Vorhergehenden für die verschiedenen Koordinatensysteme besondere Namen eingeführt, durch welche ihre verschiedene Relation zu der physikalischen Realität, dem felderzeugenden System, gekennzeichnet wird. Aus dem nämlichen Grunde wird das Wort »Feld« nur als Bezeichnung eines in einer gewissen Raumgegend von einem realen felderzeugenden Systeme erzeugten Zustandes gebraucht werden. Mit den Namen homogenes Neutralfeld, Inertialfeld oder gravitationsfreies Feld bezeichnen wir stets eine andere Realität als ein homogenes Gravitationsfeld. Dies wollen wir auch »ein pseudohomogenes Neutralfeld« nennen, um dadurch gleichzeitig seinem realen Unterschied von und seiner

phänomenologischen Gleichheit mit einem homogenen Neutralfeld einen Ausdruck zu verleihen. In dem homogenen Felde kann man sowohl ruhende oder gleichförmig bewegte Inertialsysteme als akzelerierte »Gravitationssysteme« (d. i. Koordinatensysteme, in bezug auf welche isolierte Körper Beschleunigung haben) und in dem pseudohomogenen sowohl ruhende oder gleichförmig bewegte (oder akzelerierte) Gravitationssysteme als feldakzelerierte Inertialsysteme wählen. Zufolge des zweiten Aequivalenzprinzips lassen sich alle physikalischen Vorgänge innerhalb eines materiellen Systems S in dem ersten Felde und eines materiellen Systems S_1 in dem zweiten Felde in ganz gleicher Weise beschreiben, wenn wir für beide ein Inertialsystem oder für beide ein passendes Gravitationssystem zur Beschreibung wählen; aber wenn wir ausser den Systemen S und S_1 die felderzeugenden Systeme in die Betrachtung mit hineinziehen, so werden wir erkennen, dass der Unterschied ein durchaus realer ist und u. a. bei der Koordinatenwahl dadurch hervortritt, dass die Inertialsysteme in dem ersten Feld gleichförmig bewegt, in dem zweiten relativ zu dem felderzeugenden Systeme (oder in dessen Neutralfeld) feldakzeleriert sein müssen.

Aber ebenso wenig wie die Koordinatensysteme, welche das zweite Aequivalenzprinzip als äquivalent aufstellen, schlechthin als in physikalischer Hinsicht gleichgestellt angesehen werden können, ebenso wenig ist dies der Fall mit denen, welche das erste als äquivalent aufstellen. Unter den verschiedenen relativ zueinander gleichförmig bewegten Inertialsystemen ist tatsächlich das primäre, das in dem Neutralfelde in Ruhe ist, durch seine besondere Relation zu dem felderzeugenden Systeme — also zu einem Systeme von prinzipiell erkennbaren Körpern, die für den Zustand

in der betreffenden Raumgegend bestimmend sind — vor den übrigen bevorzugt. In der ursprünglichen, speziellen Relativitätstheorie war diese äussere Realität überhaupt nicht ins Auge gefasst; man hatte nur mit den Inertialsystemen selbst zu tun, und sie wurden alle als vollkommen gleichgestellt angesehen. Darum war die Einordnung des Beobachteten (Ereignisse und materielle Punkte) nach Zeit und Raum, welche ein Beobachter in einem von den Systemen unternahm, als vollständig gleichgestellt mit derjenigen anzusehen, die ein Beobachter in jedem andern von den Systemen nach den nämlichen Regeln unternahm, und es wurde als eine tiefgehende erkenntnistheoretische Entdeckung aufgefasst, dass die Urteile der verschiedenen Beobachter über Längen und Zeiten sich nicht vereinbaren liessen. J. Petzoldt hat in seiner früher zitierten Abhandlung ein ausdrucksvolles und treffendes Bild von der scheinbar vorliegenden Situation gegeben, indem er sagt, dass jeder von den relativ zueinander bewegten Beobachtern als eine Monade in seiner eigenen Raum-Zeitwelt lebt, ohne Fenster zu den Welten der andern. Aber das felderzeugende System, das in der allgemeinen Relativitätstheorie in den Formeln implicite zugegen ist (durch die zehn Komponenten des Gravitationspotentials vertreten), und das in der speziellen nur »unsichtbar« geworden ist, eröffnet jedem von den Beobachtern Aussicht in die Welten der andern.

Das felderzeugende System, das aus erkennbaren Körpern besteht, liefert die erforderliche reale Basis für eine kausale Betrachtung von den Aenderungen der festen Körper und der Perioden der periodischen Systeme. Dass eine Geschwindigkeitsvergrösserung im Neutralfelde z. B. stets einen langsameren Gang einer Uhr mit sich führt,

während eine Geschwindigkeitsverminderung im Neutralfelde stets einen schnelleren Gang zur Folge hat (vergl. die Fussnote S. 5), streitet nicht wider die kausale Relativitätsforderung, da es eine gesetzmässige Relation zwischen prinzipiell beobachtbaren Dingen ist. — Es muss hervorgehoben werden, dass für die Grundfrage von der Möglichkeit einer kausalen Klarlegung von den Aenderungen nicht die besondere elektrodynamische Erklärung, sondern nur die Bevorzugung eines von den Inertialsystemen von Belang ist.

Wenn jemand die Frage stellen würde, ob die Bewegungen im Neutralfelde wirklich solche Aenderungen von Dimensionen und Periodenlängen hervorbringen, so lässt sich darauf antworten, dass wir diese Aenderungen in dem nämlichen Sinne wie alle anderen Aenderungen von Längen und Zeiten (z. B. Aenderungen von Masstablängen und Uhrperioden durch Druck und Temperatur), welche durch beobachtbare Ursachen bestimmt sind, als wirklich auffassen können. Wir brauchen gar nicht dem Wort wirklich — ebenso wenig wie dem Wort Ursache — eine gewisse transzendente Bedeutung zuzuschreiben. Das, wovon es sich handelt, ist in diesem Falle, wie in allen anderen Fällen, nur, ob die Aenderungen sich in gesetzmässige eindeutige Relationen zu beobachtbaren Dingen bringen lassen, oder anders ausgedrückt, ob sie in einer eindeutigen (d. h. für alle Beobachter geltenden) kausalen Beschreibung ihren Platz finden. Wir können in diesem Zusammenhang daran erinnern, dass wir bei unseren Messungen überhaupt nicht Längen und Zeiten unmittelbar beobachten, sondern, wie Einstein hervorgehoben hat, nur raumzeitliche Koinzidenzen konstatieren.

Allerdings können wir hier von der Erde aus nicht alle Körper, die das Feld erzeugen, beobachten, aber sie sind

prinzipiell beobachtbar, und einen Teil von ihnen, nämlich die sichtbaren Sterne können wir in der Tat beobachten. Mittels lange Zeiten hindurch fortgesetzter Beobachtungen können wir zudem einigen Aufschluss über die Bewegung des Sonnensystems relativ zu diesen Sternen erlangen, und wenn man sie in dieser Hinsicht als vollgültige Vertreter der ganzen für unsere Raumgegenden bestimmenden felderzeugenden Welt ansehen darf, könnte man auf diesem Weg das primäre Inertialsystem oder, anders ausgedrückt, die Geschwindigkeit der Erde und der irdischen Körper im Neutralfelde bestimmen. Nun können wir wohl nicht davon ausgehen, dass die genannte Voraussetzung richtig ist; denn das sichtbare Sternensystem ist vielleicht nur ein sehr geringer Teil des ganzen felderzeugenden Systems (in der Tat müssen wir dies annehmen), und es braucht relativ zu dem diesen repräsentierenden »Ersatzkörper« nicht in Ruhe zu sein. Aber was hier von Belang ist, ist nur, dass es prinzipiell möglich ist, die Bewegung in dem von diesem erzeugten Neutralfelde zu bestimmen. Wenn wir diese Bewegung kännten, so könnten wir unter Anwendung der Lorentz'schen Theorie — mit dem Neutralfelde anstatt des Weltäthers — die der Geschwindigkeit der betreffenden Körper entsprechenden Verkürzungen und Periodenvergrösserungen bestimmen; jeder Beobachter könnte dann seine Masstäbe und Uhren für diese korrigieren und mittels dieser Korrektionen eine Raum-Zeit-Einordnung von allen Beobachtungen gewinnen, welche mit der des »primären Beobachters« stimmen würde.

Man kann diese Einordnung als kausal bezeichnen, weil sie die Raum-Zeit-Bestimmung auf einem gesetzmässigen Zusammenhang zwischen den Längen der Masstäbe beziehungsweise der Periodenlängen der Uhren und der Rela-

tum der Massstäbe beziehungsweise der Uhren zu realen Körpern (und deren physikalischen Eigenschaften) basiert, während die Einstein'sche Einordnung nicht kausal, sondern definitionsmässig ist. Da Einsteins Gleichzeitigkeitsdefinition sich auf das gewählte Inertialsystem bezieht und der Geschwindigkeitsbegriff mit dem Gleichzeitigkeitsbegriff untrennbar verknüpft ist, wird auch die Einstein'sche »Lichtgeschwindigkeit« nur eine Definitionsgrösse sein, in Relation zu dem zufälligen Inertialsystem, das an sich nichts mit der Lichtbewegung zu schaffen hat. Es ist daher nicht — wie es manchem scheint — eine Absurdität, dass diese Lichtgeschwindigkeit in den verschiedenen Inertialsystemen dieselbe ist: das steht gar nicht, wie gewöhnlich behauptet wird, in einem entscheidenden Widerspruch mit der gewöhnlichen Raum- und Zeitauffassung. Nach dem Einstein'schen Einordnungsprinzip muss man auch die Länge eines Einheitsmasstabes und die Periode einer Normaluhr, welche nebst dem Beobachter in einem gewissen Inertialsysteme ruhen, als unverändert (gleich der Einheit) rechnen, wenn man ihnen und dem Beobachter eine gewisse Geschwindigkeit erteilt und dann ein neues Inertialsystem, in welchem sie jetzt ruhen, wählt; bei der kausalen Einordnung muss man dagegen — wegen der veränderten Geschwindigkeit im Neutralfelde — eine Korrektion einführen. Aber die beiden Einordnungsprinzipien stehen nicht in realem Widerspruch miteinander; das Einstein'sche kann als eine mathematische Transformation von dem kausalen betrachtet werden — und noch dazu eine berechtigte und wertvolle Transformation, weil sie grosse Einfachheit in der Beschreibung mit sich führt.

Es könnte den Anschein haben, dass Einstein selbst der im Vorhergehenden entwickelten Auffassung, nach

welcher das von Massen im Raume erzeugte Feld den früheren Weltäther ersetzt, jetzt nicht fern steht. Wir können in dieser Hinsicht auf einen Artikel von Einstein in »Die Naturwissenschaften«, Jahrg. 1918 S. 697, hinweisen. Er sagt hier u. a., dass man nicht einen Raum ohne Materie (elektromagnetische Felder) als schlechthin leer zu betrachten braucht; nach der allgemeinen Relativitätstheorie hat er physikalische Qualitäten, welche durch die Komponenten des Gravitationspotentials mathematisch charakterisiert sind, und nach dieser Auffassung ist es zulässig, »dass man von einem Aether spricht, dessen Zustand von Punkt zu Punkt stetig variiert«. Und wie früher erwähnt, hat er »ferne Massen als die realen Ursachen des Unterschieds zwischen »galileischen« und anderen Koordinatensystemen erkannt. Er zieht aber nicht die, wie mir scheint, unvermeidliche Folgerung, dass sie dann auch die vollkommne Gleichstellung der verschiedenen gegeneinander gleichförmig bewegten Inertialsysteme aufheben. Im Gegenteil scheint es, dass er einem solchen Gedanken gegenüber Front macht. An der zitierten Stelle sagt er u. a.: »Gemäss der speziellen Relativitätstheorie gab es keinen bevorzugten Bewegungszustand mehr; dies bedeutet Leugnung des Aethers im Sinne der früheren Theorien. Denn gab es einen Aether, so musste er in jedem Raum-Zeitpunkt einen bestimmten Bewegungszustand haben, der in der Optik eine Rolle spielen musste. Einen solchen bevorzugten Bewegungszustand aber gibt es nicht, wie die spezielle Relativitätstheorie lehrte, und darum gibt es auch keinen Aether im alten Sinne«. Er warnt zudem ausdrücklich davor, dem »Aether«, welchen er anerkennen kann, »an jeder Stelle eine bestimmte Geschwindigkeit zuzuschreiben«. Insofern dies so zu verstehen ist, dass man nicht dem »Aether«

und dem betrachteten materiellen System eine bestimmte Geschwindigkeit relativ zueinander zuschreiben darf, fordert er, scheint es mir, eben in einem sehr entscheidenden Punkte ohne irgend eine im Wesen der Relativitätstheorie begründete Notwendigkeit das Festhalten an einer akausalen Auffassung. Dass es nicht nur der Gedanke von einem besonderen über die ganze Welt gleichartigen Aethermedium ist, welchen er abweist, geht u. a. aus den (von mir) hervorgehobenen Worten in dem Zitate hervor. Denn die Forderung, dass ein bestimmter Bewegungszustand relativ zu dem Aether eine Rolle in der Optik spielen, oder mit andern Worten in irgend einer Weise in den physikalischen Vorgängen hervortreten muss, fällt, wie es scheint, nicht weg, weil man das Wort Aether im Sinne von dem Neutralfelde in einer Raumgegend, wo dieses homogen (oder pseudohomogen) ist, anwendet. Sie steht unzweifelhaft eher da als ein Ausdruck von Einsteins Auffassung des Verhältnisses zwischen dem Phänomenologischen und dem Realen. Nur was phänomenologisch hervortritt, existiert, könnte man sagen. Wenn die verschiedenen gegeneinander gleichförmig bewegten Inertialsysteme wirklich phänomenologisch äquivalent sind, sollte es dann unberechtigt sein, eine Erklärung oder »kausale Beschreibung« auf einer Bevorzugung von einem unter ihnen zu basieren, und die nämliche Betrachtung könnte man gegenüber dem Verhältnis zwischen feldakzelerierten und unakzelerierten Inertialsystemen anstellen. Da es dennoch unzweifelhaft ist, dass gewisse Inertialsysteme sich vor andern durch ihre Relation zu dem felderzeugenden System auszeichnen, scheint es, als ob hier ein Widerspruch zwischen dem Aequivalenzgesichtspunkte und der entwickelten kausalen Betrachtungsweise besteht.

Der Widerspruch löst sich doch auf, wenn man in Betracht zieht, dass sowohl ein homogenes als ein pseudohomogenes Neutralfeld nur ein idealer Grenzfall ist. Es ist zwar richtig, dass die gleichförmige Bewegung von einem sich selbst überlassenen materiellen Systeme in einem homogenen Neutralfelde oder sein freier Fall in einem pseudohomogenen Neutralfelde der Relativitätstheorie zufolge sich an keinen Vorgängen innerhalb der Grenzen des Systems erkennen lässt; aber die Bedingung für das vollkommene Verschwinden von jedem solchen Kriterium der Bewegung ist freilich, dass das Feld in der betreffenden Raumgegend vollkommen homogen oder pseudohomogen ist. In der Tat kann aber diese Bedingung für Raumgegenden von endlicher Ausdehnung nicht erfüllt sein. (Nur für eine infinitesimale Raumgegend kann man durch passende Koordinatenwahl erreichen, dass die Einstein'schen Komponenten des Gravitationspotentials die einem Inertialsysteme entsprechenden speziellen Werte annehmen). Das phänomenologische Kriterium einer Bewegung im Aether (d. i. dem Neutralfelde), welches Einstein fordert, ist in der Tat immer zugegen für ein endliches materielles System und bei genügend feinen Beobachtungsmitteln. Dass es an der Grenze verschwindet, d. i., wenn die Dimensionen des materiellen Systems Null werden oder der Raumzustand einen idealen Grenzzustand erreicht, kann man nicht mit Recht als Argument dafür gebrauchen, dass man nicht Bewegung relativ zum Neutralfelde als physikalische Ursache einführen darf. In der Physik wie in der Mathematik darf man nur mit grosser Vorsicht von Idealfällen aus räsonnieren.

Betrachten wir das zu einem einzelnen Weltkörper gehörige Neutralfeld, leuchtet es unmittelbar ein, dass es

nicht homogen sein kann. Wenn wir voraussetzen, dass die Beschleunigung gegen den Weltkörper eine Funktion der Entfernung von seinem Zentrum ist, sieht man, dass das Feld für eine endliche Raumgegend auch nicht pseudohomogen sein kann, denn es ist ja eben die konstante Beschleunigung, die ein pseudohomogenes Neutralfeld oder ein homogenes Gravitationsfeld kennzeichnet. Dagegen könnte es scheinen, dass eine Möglichkeit dafür bestehe, dass eine nicht sphärische Masse ein homogenes oder pseudohomogenes Neutralfeld von endlicher Ausdehnung bilden könne. Wenn die Feldintensität wie das Newton'sche Gravitationspotential bestimmt würde, dann wäre sie ja konstant im Raume innerhalb einer homogenen Kugelschale. In der Tat ist es doch ausgeschlossen, dass ein solcher vollkommen homogener (oder pseudohomogener) Zustand existieren kann, schon weil das Feld von individuellen Körpern — Weltkörpern oder Atomen — erzeugt ist. Hieraus folgt, dass es eine von diesen bestimmte Struktur besitzt, und diese Eigenschaft des »Neutralfeldäthers« ist von grossem Belang mit Rücksicht auf seine erkenntnistheoretische Berechtigung, als Basis für eine kausale Beschreibung des Weltgeschehens zu dienen. Wenn er ein vollkommen unterschiedsloses Kontinuum wäre, könnte man nicht wohl Bewegungen in ihm als wahre physikalische Ursachen der Bildung von magnetischen Feldern, Dimensionsänderungen u. s. w. ansehen. Man könnte allerdings auch in diesem gedachten Falle die Bewegungen relativ zu dem Stoffabschnitte des felderzeugenden Systems als bestimmende Faktoren in die physikalische Beschreibung einführen, und wenn man also die Vorgänge in gesetzmässige Relationen zu beobachtbaren Dingen bringt, hat man die kausale Relativitätsforderung erfüllt, aber man könnte es doch als

eine Rückkehr zu einer Fernwirkungsauffassung schroffer Art bezeichnen, wenn die Eigentümlichkeiten des fernen Systems sich nicht in dem Felde mittels Variationen — wie schwach sie auch wären — in seinem Zustande abspiegelten und jeder kleine endliche Teil des Feldes dadurch ein Abbild der ganzen äusseren Welt würde.

Es liegt Grund vor, hier hervorzuheben, dass der Raumzustand, der mit dem Namen »Neutralfeldäther« bezeichnet wird, ein Kombinationszustand ist, nicht nur in dem Sinne, dass das ganze Neutralfeld aus vielen individuellen Neutralfeldern (denen der einzelnen Atome) zusammengesetzt ist, sondern dass das einzelne Atomneutralfeld für sich als ein Spezialzustand angesehen werden muss, welcher durch spezielle Kombinationen von primitiveren Elementen — den positiven und negativen Grundbestandteilen — zu neutralen Atomen bedingt ist. Wenn es möglich ist, uns alle physikalischen Vorgänge als in dem Neutralfelde verlaufend vorzustellen und in der Relativitätstheorie diesen »Aether« als für »die metrischen Eigenschaften des Raumes« bestimmend aufzufassen, kommt das daher, dass die enorme Menge von neutralem Stoffe im wesentlichen den Charakter des Raumzustandes bestimmt. Denken wir uns aber einen Raum, worin es z. B. nur negative Elektronen gibt, so haben wir hier einen »Aether« ganz anderer Art. Man könnte wohl auch fragen, ob die Abweichungen von der klassischen Elektrodynamik, welche durch die Entwicklung der Atomtheorie zu Tage gebracht sind, vielleicht damit in Zusammenhang ständen, dass »der Aether« im Innern der Atome wegen des überwältigen Einflusses der elektrischen Atombestandteile kein Neutralfeldäther sei. — Wir wollen uns nicht in die Probleme vertiefen, zu denen diese Betrachtungen uns führen könnten; es soll hier nur

betont werden, dass man nicht davon absehen kann, wenn man den mit der Relativitätstheorie verknüpften erkenntnistheoretischen Fragen auf den Grund kommen will.

Wenn ein materielles System von endlicher Ausdehnung sich in einem willkürlich variierenden Neutralfelde bewegt, wird diese Bewegung in Krümmungen von Lichtstrahlen, Spektralverschiebungen bei Vergleichungen von Licht aus den verschiedenen Teilen des Systems u. a. m. phänomenologisch hervortreten. In der Relativitätstheorie kann man von allen diesen Verhältnissen Rechenschaft ablegen, indem man den Komponenten des Gravitationspotentials in jedem Raum-Zeitpunkte von dem vierdimensionalen Kontinuum gewisse Werte zuschreibt. Um das Beobachtete eindeutig als Ausdruck eines bestimmten Bewegungszustandes des Systems relativ zu einem Systeme von felderzeugenden Körpern auslegen zu können, muss man aber eine vollkommne Kenntnis der Gesetze besitzen, nach welchen der Zustand des Feldes durch die felderzeugenden Körper bestimmt wird. Denken wir uns einen Beobachter im Besitze einer solchen vollkommnen Kenntnis dieser »Neutralfeldgesetze« und zudem im Besitze einer genügenden Fülle von Beobachtungen — sowohl Messungen innerhalb des für seine Masstäbe und Uhren zugänglichen materiellen Systems als »Lichtbotschaften« von den fernen felderzeugenden Körpern — müsste es ihm prinzipiell möglich sein, alles Beobachtete in Raum und Zeit in Uebereinstimmung mit der Einordnung anderer ebenso vollkommner Beobachter einzuordnen. Aber die Bedingung für das Erlangen von einer solchen Uebereinstimmung ist die, dass alle die Beobachter sich beständig die Masstäbe und Uhren für die Variationen und Bewegungen im Felde korrigiert denken. Man könnte z. B. den Atomabstand in dem Krystallgitter

des Steinsalzes als Längeneinheit und die Periode der Lichtschwingungen der roten Kadmiumlinie an einer bestimmten Stelle der Erde und zu einem bestimmten Zeitpunkte als Zeiteinheit wählen; aber der irdische Beobachter müsste dann stets Korrektionen für die Aenderungen dieses Abstandes und dieser Periode durch Variationen und Bewegungen im Neutralfelde einführen.

Es würden sich dann für die Lichtgeschwindigkeit sowohl relativ zu dem Neutralfelde als relativ zu einem Koordinatensystem, worin er in Ruhe wäre, an verschiedenen Orten und zu verschiedenen Zeitpunkten verschiedene Werte finden. Wenn dagegen der genannte Abstand und die genannte Periode schlechthin als Einheiten festgestellt würden, so dass jeder Beobachter ihnen stets — ohne die erwähnten Korrektionen einzuführen — den Zahlenwert 1 beilegte, wenn die Träger dieser Einheiten relativ zu ihm und in seiner unmittelbaren Nähe ruhten, dann würde er hier für die Lichtgeschwindigkeit in einem relativ zu ihm ruhenden Koordinatensystem der Relativitätstheorie zufolge stets den nämlichen Wert finden. Aber die Einordnungen verschiedener Beobachter würden sich dann voneinander unterscheiden. Unter der nämlichen Voraussetzung wird man bei der Messung von festen Körpern im allgemeinen nicht Uebereinstimmung mit den Regeln der euklidischen Geometrie finden. In der allgemeinen Relativitätstheorie wird dies gern in der Weise ausgedrückt, dass die euklidische Geometrie in dem betreffenden Raum nicht gilt. Wenn man aber für den Einfluss der Feldvariationen und Bewegungen auf den Masstab und den gemessenen Körper ganz wie für den Einfluss von Temperatur und Druck korrigiert, sollte man natürlich erwarten, dass die Abweichungen von den euklidischen Regeln verschwinden würden — ebenso wie

die Abweichungen, welche man z. B. bei einer festen Scheibe findet, wenn die Temperatur an der Mitte höher ist als am Rande[1]. Wenn man durch solche Korrektionen die Messresultate auf die euklidischen Regeln zurückführen kann, unter Anwendung von Neutralfeldgesetzen, die nicht wider die von der allgemeinen Relativitätstheorie abgeleiteten Relationen streiten, dann ist es möglich, die allgemeine Relativitätstheorie wie die spezielle als eine mathematische Transformation einer auf dem Grunde der gewöhnlichen Raum- und Zeitauffassung ruhenden Kausaltheorie zu betrachten.

Es könnte vielleicht so scheinen, als ob diese Möglichkeit von vorn herein durch mehrere von Einstein und andern angestellte Betrachtungen abgewiesen wird.

Einstein denkt sich z. B. (»Über die spez. u. die allg. Relativitätstheorie« S. 55), dass die Peripherie und der Radius einer rotierenden Scheibe mit einem Einheitsmasstab (einem relativ zum Radius kleinen Stab) von einem Beobachter auf der Scheibe gemessen werden; dann findet dieser — wenn er mit hinlänglicher Feinheit messen könnte — für das Verhältnis zwischen Peripherie und Radius einen grössern Zahlenwert als 2π, weil der Stab eine Lorentz-Kontraktion erfährt, wenn er an den Rand gelegt wird und dadurch in seiner eigenen Richtung in dem Inertialsystem bewegt wird, in dem die Scheibe sich dreht. Hierüber sagt Einstein dann: »Damit ist bereits bewiesen, dass die Sätze der euklidischen Geometrie auf der rotierenden Scheibe und damit überhaupt in einem Gravitationsfelde nicht

[1] Selbst wenn der benutzte Masstab aus dem nämlichen Stoff wie die Scheibe ist und er überall die Temperatur des Ortes erlangt, wird man durch Messung der Peripherie und des Radius für ihr Verhältnis einen grösseren Zahlenwert als 2π finden, wenn die erforderlichen Korrektionen nicht eingeführt werden.

genau gelten können, wenigstens wenn man dem Stäbchen überall und in jeder Orientierung die Länge 1 zuschreibt. In dem letzten Satze ist aber ein für die kausale Betrachtung entscheidender Vorbehalt genommen, denn bei der kausalen Einordnung dürfen wir ja eben dieses nicht tun, sondern müssen im Gegenteil die Stablänge für die Lorentz-Verkürzung korrigieren, und dadurch kommt in dem betrachteten Fall eben der euklidische Wert 2π für das Zahlenverhältnis heraus. Dass Einstein sich dieser Betrachtungsweise nicht anschliesst, liegt offenbar darin, dass er ein mit dem Beobachter auf der rotierenden Scheibe folgendes Koordinatensystem — in welchem der Stab keine Bewegung hat — als vollkommen gleichberechtigt mit dem Inertialsystem betrachtet. Aber die Rotation der Scheibe ist doch eine Bewegung relativ zu dem äussern felderzeugenden System, und darin haben wir eine reale Basis für eine Bevorzugung des Inertialsystems vor dem rotierenden Koordinatensystem und für die kausale Auffassung der Kontraktion des Stabes — der Bevorzugung des primären Inertialsystems in der speziellen Relativitätstheorie analog.

In ähnlicher Weise verhält es sich mit den mehr allgemeinen Betrachtungen, durch die man dartun kann, dass die Kovarianzforderung sich unmöglich erfüllen lässt, wenn man nicht dem Ausdruck eines Linienelements eine nichteuklidische Form erteilt (Siehe z. B. E. Freundlich: Die Grundlagen der Einstein'schen Gravitationstheorie 1916 S. 31). Dieser Satz ist sicherlich richtig, aber daraus, dass man sich, um die Kovarianzforderung zu erfüllen, von der euklidischen Geometrie entfernen muss, darf man nicht schliessen, dass die Möglichkeit einer mit der euklidischen Geometrie stimmenden kausalen Einordnung mit Einsteins Relativitätstheorie unvereinbar ist. Denn die kausale Re-

lativitätsforderung führt nicht mit sich, dass man die physikalischen Gesetze in kovarianter Form ausdrücken soll, sondern nur, dass die Koordinatensysteme, in bezug auf welche Gesetze ausgedrückt werden, andernfalls gewisse nicht mit einbezogene Realitäten vertreten; aber andererseits schliesst sie nicht aus, dass man durch ein besonderes mathematisches Verfahren die Gesetze in kovarianter Form ausdrücken kann (vergl. S. 12 u. 15). Man kann sagen, dass bei einer Kausaltheorie (d. i. einer Darstellung, in welcher der kausale Zusammenhang explicite hervortritt) die von den Massen im Raume bedingten Feldzustände in die Reihe der übrigen für die physikalischen Vorgänge (und die Dimensionen der Körper) bestimmenden Ursachen eingehen sollen. In der Einstein'schen Relativitätstheorie gibt man aber der »Gravitation« eine Sonderstellung als bestimmend für »die metrischen Eigenschaften des Raumes«. Durch diese — überaus fruchtbare — mathematische Transformation wird das Kausale in diesem Punkte verschleiert, aber es wird nicht aus der Welt gebracht.

Wie tief diese Sonderstellung der »Gravitation« (d. i. des Neutralfeldzustandes) im Wesen des Geschehens begründet ist, soll hier nicht näher erörtert werden. Wir wollen nur auf die Bemerkungen hinweisen, welche S. 45 von dem »Neutralfeldäther« und S. 33 von einer etwaigen Ausnahmestellung der Quantenphänomene gemacht wurden — und endlich auch darauf, dass die Sonderstellung eine momentane Anpassung von Masstäben und Uhren nach dem Zustande im betreffenden Raum-Zeitpunkte (den Gravitationspotentialen) involviert, während es nach der kausalen Auffassung eine sehr naheliegende Annahme ist, dass diese Anpassung eine gewisse Zeit erfordert, — wie z. B. Anpassungen nach der Temperatur eines Milieus.

V.

Im Vorhergehenden haben wir von dem felderzeugenden System gesprochen und darunter die ganze äussere materielle Welt verstanden, die zur Bildung des Neutralfeldes in der betreffenden Raumgegend beiträgt. Es ist nun eine Frage von grossem Interesse mit Rücksicht auf die kausale Relativitätsforderung, ob es möglich ist, dem materiellen Universum unendliche Ausdehnung und Masse zuzuschreiben, und ob man in diesem Falle zu der für eine gewisse Raumgegend bestimmenden felderzeugenden Welt das ganze Universum mitrechnen muss, oder ob man sich damit begnügen kann, einen begrenzten endlichen Teil in Betracht zu ziehen, indem der Einfluss fernerer Weltgegenden verschwindend gering wird.

Es ist wohlbekannt, dass die Annahme einer unendlichen Welt, wo die durchschnittliche Dichte für hinlänglich grosse Räume überall einen endlichen Wert hat, mit Newtons Theorie unvereinbar ist. Betrachten wir nämlich eine Kugel mit Radius r und Dichte ρ, haben wir für das Potential im Zentrum den Wert $2\pi\rho r^2$ (wenn wir von der Gravitationskonstante absehen), und wenn ρ endlich ist, wird das Potential mit dem Radius gegen Unendlich hin wachsen. Hieraus folgt, dass das Potential in jedem Punkte einer solchen Welt den Wert ∞ annimmt, und damit wird die Kraft unbestimmt.

Denken wir uns dagegen, das ρ in stärkerem Verhältnis als $\frac{1}{r^2}$ gegen Null hin abnimmt, wenn wir aus unserer Raumgegend nach dem Unendlichen hin gehen, wird das Potential endlich sein. Wenn die gesamte Masse aber auch trotzdem unendlich gross sein kann, können wir

eine Welt dieser Art als »eine endliche Mittelpunktswelt« bezeichnen. Sowohl die einzelnen Weltkörper als die Strahlenenergie wird aber nach und nach ein solches Weltsystem verlassen und sich in dem unendlichen Weltraume verteilen. Aber selbst hiervon abgesehen, ist eine Weltvorstellung dieser Art wenig befriedigend für unsern Gedanken.

Eine Weltordnung, bei welcher ρ nicht mit wachsendem Abstande gegen Null hin abnimmt, wäre zulässig, wenn man in Newtons Gravitationsgesetz eine geeignete kleine Aenderung einführte. Es würde nichts helfen, den Exponenten des Abstandes r ein wenig von 2 abweichen zu lassen; dagegen könnte man z. B. den Zweck dadurch erreichen, dass man in dem Gesetze ausser der Gravitationskonstante einen Faktor $e^{\lambda r}$ einführte, wo λ eine kleine Zahl ist — oder wenn man, wie man es auch ausdrücken kann, die Gravitation eine »Absorption« im Raume erfahren liesse.

Die Vorstellung eines unendlich ausgedehnten materiellen Universums ohne Mittelpunkt lässt sich aber auch mit der unmodifizierten Newton-Theorie vereinbaren, indem wir unsere Vorstellung von dessen Bau auf die Ordnung des Stoffes in kleineren und grösseren Aufhäufungen basieren, welche uns, so weit unsere Beobachtungen reichen, als eine Tatsache begegnen. Wir wissen, dass die Atome im wesentlichen in Welkörpern wie Sonnen, Planeten u. a. m. gesammelt sind. Diese Weltkörper bilden Sonnensysteme, die in Sternenhaufen gesammelt sind, welche wiederum — vielleicht mit Aufhäufungen von dazwischenliegenden Grössenordnungen als Mittelgliedern — das System der Milchstrasse bilden. Wir haben Grund dazu, gewisse Spiralnebel, welche sich unter anderm durch grössere Geschwindigkeiten als die für die Sterne des Milchstrassensystems durch-

schnittliche auszeichnen, für Sternensysteme ähnlicher Art wie unser Milchstrassensystem anzusehen. Länger als zu diesen Systemen reichen unsere Beobachtungen nicht; das findet aber eine durchaus hinlängliche Erklärung in der Beschränkung unserer Beobachtungsmittel, und wir haben keinen Grund, die Erklärung in einer Weltbegrenzung zu suchen. Es ist vielmehr eine naheliegende und wahrscheinliche Annahme, dass es unter diesen Systemen eine sehr grosse Anzahl gibt, welche zusammen ein endliches System höherer Grössenordnung bilden, und dass diese Grossysteme wiederum ein endliches System noch höherer Grössenordnung bilden. Wenn wir uns den Stoff in dem Weltraume in Systeme von ins Unendliche wachsender Grössenordnung gesammelt denken, folgt unser Gedanke dann nur der von unsern Beobachtungen angezeigten Linie immer weiter in der eingeschlagenen Richtung — statt in beschränkter egozentrischer Weise an einem ganz zufälligen von unserm beschränkten Beobachtungsvermögen bestimmten Punkte anzuhalten.

Der Gedanke von einer solchen Ordnung der Sterne wurde im Prinzipe schon in dem 18. Jahrh. von Lambert ausgesprochen, und in der Astronomie hat man ihn jedenfalls für die Lichtverhältnisse gegenüber dem Unendlichkeitsprobleme in Anwendung gebracht. Es ist nämlich erkannt, dass das Universum sehr wohl in dieser Weise aus einer unendlich grossen Zahl von leuchtenden Sternen erbaut sein kann, ohne dass der Himmel darum durch das gesamte Licht aller dieser Sterne ein einziges Lichtmeer wird. Dass man auch durch eine Ordnung dieser Art der Konsequenz entgehen kann, dass das Newton'sche Potential in irgend einer Stelle unendlich wird, ist leicht zu sehen.

Wir wollen der Einfachheit halber annehmen, dass jedes System eine Kugel ist und aus einer sehr grossen Zahl von »Kugelsystemen» nächster unterer Grössenordnung besteht, dass diese Systeme gleichförmig verteilt sind, dass sie eine sehr geringe Masse relativ zu der des grössern Systems besitzen, und dass ihre Radien relativ zu ihren gegenseitigen Abständen sehr klein sind. Wir denken uns vorläufig nur zwei Grössenordnungen, nämlich ein System S_2 mit Durchschnittsdichte ρ_2 und Radius r_2 und als seine Komponenten Systeme S_1 mit den entsprechenden Werten ρ_1 und r_1. In einem Punkte innerhalb S_2 in dem Abstande a_2 von seinem Zentrum wird das Potential dann sehr annähernd $P_2 = {}^2/_3\pi\rho_2(3r_2{}^2 - a_2{}^2)$ sein, wenn der Punkt nicht irgend einem von den Systemen S_1 besonders nahe liegt. Wenn er dagegen innerhalb eines von diesen Systemen liegt, gibt dieses für sich betrachtet ein Potential P_1 von der nämlichen Form wie P_2. Da das von allen übrigen Systemen S_1 zusammen erzeugte Potential den gemachten Voraussetzungen zufolge sehr nahezu gleich P_1 ist, wird das ganze Potential annähernd gleich $P_1 + P_2$.

Wenn wir nun statt der beiden Systemordnungen S_1 und S_2 eine unendliche Reihe S_1, S_2, S_3 . . . . . haben, von welchen jede Systemordnung in der nämlichen Relation zu der folgenden wie S_1 und S_2 steht, können wir das Potential P in einem Punkte von einem von den Systemen der untersten Grössenordnung S_1 annähernd als eine Summe von Gliedern ausdrücken:

$$P = P_1 + P_2 + \ldots = {}^2/_3\pi\rho_1(3r_1{}^2 - a_1{}^2) + {}^2/_3\pi\rho_2(3r_2 - a_2{}^2) + \ldots$$

Wenn P einen endlichen Wert annehmen soll, ist es nicht hinlänglich, dass $\rho_n r_n{}^2$ sich Null nähert, indem n gegen unendlich hin wächst. Dass Produkt muss sich Null

in solcher Weise nähern, dass die Reihe gegen einen endlichen Wert hin konvergiert. Wenn diese Bedingung erfüllt ist und man so viele Glieder mitnimmt, dass man sich dadurch diesem Grenzwert genügend nähert, kann man die übrigen Glieder weglassen oder, anders ausgedrückt, die Systeme höherer Grössenordnung als das letzte mitgenommene Glied haben nur einen überaus kleinen Einfluss auf die Verhältnisse in der betreffenden Raumgegend. Nur ein endlicher Teil von dem Weltall hat Bedeutung für sie; »ihre physikalische Welt« ist endlich.

Für einen Punkt, der innerhalb der Grenzen eines gewissen Systems S_n liegt, aber weder einem System S_{n-1} angehört noch in der Nähe von einem dieser Systeme liegt, werden alle Glieder von P links von P_n wegfallen. Körper, die sich in solchen Punkten befinden, könnte man Vagabunden in S_n nennen (z. B. Meteore, welche in einem Sternenhaufen wandeln, ohne irgend einem von dessen Sonnensystemen so nahe zu sein, dass dessen Beitrag zu dem Potential gegenüber dem, welcher von dem Sternenhaufen als Ganzes herrührt, von Belang ist).

Es ist selbstverständlich unwesentlich, ob die Systeme Kugeln sind und die Stoffverteilung eine gleichförmige ist. Die Bedingung dafür, dass eine der Systemordnung entsprechende Auflösung von P in Glieder statt finden kann, ist nur die mehr allgemeine von der Kleinheit jedes Systems relativ zu den nächsten grösseren und zu seinen Abständen von Systemen gleicher Grössenordnung.

Dem Gravitationspotential in Newtons Theorie entspricht in unserer Auffassung von dem Weltgeschehen das, was wir die Feldintensität genannt haben, und wenn diese sich auch nicht ohne weiteres mit dem Newton'schen Potential identifizieren lässt, stossen wir doch hier auf die nämliche

Unendlichkeitsschwierigkeit, und eine Ordnung des Stoffes im Weltall in Systeme von wachsender Grössenordnung kann uns auch hier Hilfe leisten. Eben eine solche Vorstellung von dem Bau des Weltalls harmoniert zudem besonders schön mit der Auffassung von dem Weltgeshehen, welche der kausalen Relativitätsforderung entspriesst. Nach dieser Vorstellung haben wir volle Symmetrie in dem ganzen Weltall. In den fernsten Gegenden des Universums gibt es mächtige Sternensysteme gleicher Art wie die uns bekannten, aber mit andern felderzeugenden Welten als der unsrigen. Ruhekoordinatensysteme für Gegenden von solchen verschiedenen Welten können in jeder Translation und Rotation relativ zueinander sein. Ihre Lagen und Bewegungen können wir auf jedes ganz zufällige rechtwinkelige Koordinatensystem beziehen, welches wir uns über das ganze Universum ausgedehnt denken; und alle solche Koordinatensysteme sind mit Rücksicht auf die Einordnung von dem Geschehen gleichgestellt. Einfache Ausdrücke für die physikalischen Gesetze bekommen wir nicht in bezug auf ein solches universales Koordinatensystem. Aber der Umstand, dass jedes von den Systemen des Weltbaus — von den kleinsten Weltkörpern und aufwärts in die Reihe von Grössenordnungen — in der Regel relativ zu den Abständen von Systemen ihresgleichen sehr klein ist, ist von grosser Bedeutung für die Ausnutzung der Einfachheit, welche die Aequivalenzprinzipien der Relativitätstheorie repräsentieren; denn daraus resultiert, dass das ganze System sich in einem von seinen gleichen samt höheren Systemen gebildeten pseudohomogenen Neutralfelde befinden wird, und man kann dann ein Koordinatensystem so wählen, dass es ein Inertialsystem für die ganze das System umfassende Raumgegend ist.

Es wird mit diesem Weltbild harmonieren, wenn wir uns denken, dass die Durchschnittswerte von den gegenseitigen Geschwindigkeiten der gleichgestellten Systeme mit wachsender Grössenordnung zunehmen. Die Lichtgeschwindigkeit in den Räumen zwischen gleichgestellten Systemen wird aber auch mit der Grössenordnung zunehmen, da die Feldintensität abnimmt, indem wir nach und nach die kleineren Systeme verlassen. Ein Beobachter, der mit Masstab und Uhr von unserer Raumgegend nach solchen Gegenden mit kleinerer Feldintensität reiste, würde doch, wie früher gesagt, nach der Relativitätstheorie stets den gleichen Wert der Lichtgeschwindigkeit finden, wenn er nicht für die Veränderung von Masstab und Uhr korrigierte. Man könnte sich dann auch leicht denken, dass es überhaupt unmöglich wäre, durch Beobachtungen innerhalb des Sonnensystems zu merken, wenn das System nach und nach unsern Sternenhaufen, dann das Milchstrassensystem, darauf das nächste grössere System u. s. w. verlässt und Vagabund in immer grösseren Systemen wird. Die Sache ist doch nicht so einfach. Wenn die von der äusseren Welt herrührende Feldintensität kleiner wird, wird nämlich der von den Körpern des Sonnensystems gelieferte Beitrag zu der gesamten Feldintensität in jedem Punkte verhältnismässig grösser werden. Die von deren Bewegungen — z. B. der Rotation der Erde — herrührenden Störungen in dem Ruhekoordinatensysteme werden dann grösser und können merkbare Wirkungen haben (vergl. S. 27). Ebenso muss z. B. die Krümmung eines Lichtstrahls, der dicht an der Sonne vorbeipassiert, stärker werden. Man muss annehmen, dass das Verhältnis zwischen dem Krümmungsradius für einen solchen, auf den Sonnenradius winkelrechten Lichtstrahl und seinem

Abstande von dem Sonnenzentrum grösser wird, und wenn die Intensität des äusseren Neutralfeldes sich 0 nähert, kann man sich schwierig irgend einen andern Grenzwert dieses Verhältnises denken als 1, so dass die genannten Strahlen in diesem Grenzfalle einen Kreis um die Sonne beschreiben sollten. Eine andere Frage ist es, ob ein solches endliches System sich nicht schnell auflösen würde, wenn es in ein schwaches äusseres Feld gebracht würde.

Wir kommen hiermit zu der allgemeinen Frage, ob die hier geschilderte Vorstellung des Weltbaus — die ins Unendliche fortgesetze Systemordnung — uns auch gegenüber der zweiten berührten Unendlichkeitsschwierigkeit — der Zerstreuungsneigung der Strahlenenergie und der Weltkörper — helfen kann. Wir wollen doch nicht näher auf diese Frage eingehen, da sie zu dem in dieser Abhandlung behandelten Probleme nur in fernerer Relation steht. Es soll nur bemerkt werden, dass wir nicht der Zerstreuungsschwierigkeit entwischen, selbst wenn wir uns darauf beschränken, die Systemordnung zu betrachten, welche wir faktisch innerhalb des endlichen, sichtbaren Teiles der Welt vorfinden.

Dagegen wollen wir noch einen andern zur Lösung der Unendlichkeitsschwierigkeiten angewiesenen Ausweg von einem kausalen Gesichtspunkte aus betrachten. Wir meinen hiermit die Vorstellung, dass der Raum, in dem sich die physikalischen Vorgänge abspielen, »sphärisch« ist, so dass das Universum endlich sein kann, ohne Grenzen zu haben.

Dieser Ausweg ist schon bei der Newton'schen Theorie in Vorschlag gebracht worden, und Einstein hat sich dieser Auffassung angeschlossen, indem er das Weltkontinuum in bezug auf die räumliche Ausdehnung als ein nicht euklidisches dreidimensionales Volumen betrachtet, das

zwar an verschiedenen Stellen einen verschiedenen Krümmungsradius hat, aber doch im grossen und ganzen sphärisch ist (»Sitzungsber. der Kgl. Preussischen Akad. d. Wissensch.« 1917 S. 142). Diese Betrachtungsweise führt doch, wie W. de Sitter hervorgehoben hat (»Monthly Notices of The Royal Astronomical Society« LXXVIII, 1917 S. 3), zu sonderbaren und kaum annehmbaren Konsequenzen. Ausser den wirklichen Sternen müssten wir am Himmel scheinbare, »Antisterne«, sehen, welche die nämliche Klarheit wie die wirklichen und ihren Platz gegenüber diesen haben würden, wenn das Licht keine Absorption im Weltraume erfährt, und wenn die Sterne unbeweglich wären; die Bewegungen in der langen Zeit — vielleicht Millionen von Jahren — welche das Licht zu einer »Rundreise« in der Einstein'schen Welt braucht, würden doch diese Symmetrie aufheben. Einsteins Hypothese kann nach de Sitter auch nicht erklären, warum die Sterne nicht aus dem Milchstrassensysteme entwischen (l. c. S. 24).

De Sitter hat eine andere Hypothese aufgestellt. Danach ist das Weltkontinuum von sphärischem (oder elliptischem) Charakter, nicht nur in räumlicher, sondern auch in zeitlicher Hinsicht; man könnte also den Krümmungsradius von de Sitters vierdimensionalem Raum-Zeit-Kontinuum passend als den Radius einer fünfdimensionalen Hypersphäre bezeichnen. Nach de Sitters Hypothese müssten die Vorgänge in fernen Weltgegenden, von dem Gesichtspunkte des Beobachters aus beurteilt, langsamer als bei ihm verlaufen, die Spektrallinien von den fernen Spiralnebeln sollten also durchschnittlich gegen Rot verschoben sein (Die Beobachtungen geben uns keine Anzeichen von wirklicher Bedeutung in dieser Richtung). Nennen wir den Krümmungsradius R, sollte die Zeit in dem Abstand

$^1/_2 \pi R$ von dem Beobachter so zu sagen still stehen; und da das Licht deshalb unendlich langsam gehen würde, könnte es niemals zurückkehren und Antisterne bilden. Für einen Beobachter im Abstande $^1/_2 \pi R$ von uns würden natürlich nicht die Vorgänge bei ihm, sondern die Vorgänge bei uns unendlich langsam verlaufen.

Man könnte gegen die Annahme eines sphärischen (oder elliptischen) Raumes oder eines sphärischen Raum-Zeit-Kontinuums einwenden, dass eine solche Auffassung für unser Anschauungsvermögen unzugänglich ist. Wir wollen doch lieber hier die Einwendung machen, dass sie in kausaler Hinsicht nicht befriedigend ist. Was die lokalen Raumsphärizitäten (die Abweichungen von dem durchschnittlichen Krümmungsradius) betrifft, haben diese ihre kausale Basis in den durch existierende Massen bestimmten Einstein'schen Gravitationspotentialen; aber der allgemeine sphärische Charakter scheint keine solche Basis zu haben.

Was hiermit gemeint ist, wird am leichtesten verstanden werden, wenn wir die für uns anschauliche dreidimensionale Kugel betrachten und uns zweidimensionale Wesen auf der Kugelfläche lebend denken. Wir wollen annehmen, dass es für sie ebenso unmöglich ist, die dritte als für uns die vierte Dimension anzuschauen, und das deshalb die Vorstellung von einer Kugel mit Zentrum und Radius ganz inkonsistent mit ihrem Bewusstsein ist. Wenn sie aber, nachdem sie — wegen der überaus grossen Länge des Kugelradius relativ zu ihrer eigenen Ausdehnung und der ihrer beobachtbaren Umgebung — erst die Sätze der ebenen euklidischen Geometrie ausgefunden hätten, später die der sphärischen Geometrie entsprechenden Abweichungen von den euklidischen entdeckten, würden sie vorläufig physikalische Ursachen zu den beobachteten Abweichungen

suchen. Wenn es aber nicht möglich wäre, in dieser Weise einen vernünftigen Zusammenhang in dem Geschehen, d. i. einfache allgemeine physikalische Gesetze zu finden, würden einige vielleicht in der Auffassung zur Ruhe kommen, dass eine gewisse neue »Kreisgeometrie«, die ein geschlossenes logisches System bildete, und durch welche es möglich sein würde, die Vorgänge in einfachen Gesetzen zu beschreiben, die rechte, d. i. die für ihre Welt geltende, wäre. Aber in ihrer geometrisch-physikalischen Wissenschaft würde eine Linie von ganz bestimmter Länge (ein grösster Kreis auf der Kugelfläche) eine besondere Rolle spielen. Viele würden eine Erklärung der Sonderstellung dieser Länge gegenüber allen anderen Längen vermissen, und wenn sie keine beobachtbaren Realitäten als die Ursache ansehen könnten, würden sie es als einen Mangel bei dem kausalen Zusammenhang, welcher übrigens in den Vorgängen hervortritt, empfinden. Dass man, von mathematischen Analogien mit der Lehre von ebenen Kreisen geleitet, der ausgezeichneten Länge einen Namen beilegen könnte, indem man sie als 2π-mal den »Krümmungsradius des Raumes« bezeichnete, würde sie nicht befriedigen. Wir dreidimensionalen Wesen müssen zugeben, dass ihre Unzufriedenheit nicht unberechtigt wäre; wir würden reale Ursachen erkennen können, welche für den Zwang, der die Lichtstrahlen u. a. m. an der Kugelfläche festhielte, Rechenschaft ablegen könnten — Ursachen, die ausserhalb ihres Vorstellungkreises und ihrer räumlichen Welt liegen würden.

Wir werden in eine ähnliche Lage wie diese zweidimensionalen Wesen geraten, wenn wir genötigt werden, unanschauliche Analogien wie einen »sphärischen Raum« mit einem gewissen »Krümmungsradius« in unsere geometrisch-physikalische Wissenschaft einzuführen. Es wird dann in

diese ein fremdes Element, eine »unbeobachtbare Ursache« eingehen. Wenn nun die Möglichkeit, dass »ausserweltliche« Ursachen für dass Geschehen in unserer Welt mitbestimmend sind, sich auch nicht schlechthin abweisen lässt, muss man doch eine solche Annahme als in kausaler Hinsicht unbefriedigend bezeichnen.

Da der »Weltkrümmungsradius« bei der Einstein'schen Hypothese von der räumlichen Sphärizität in einer gewissen Relation zu der Gesamtmasse in der Welt steht, ist die Sachlage hier weniger klar als bei de Sitters Hypothese. Bei dieser wird die Grösse des Krümmungsradius nur durch eine gewisse Eigenschaft des Raum-Zeit-Kontinuums bedingt, und auch von den Inertialsystemen gilt, dass sie durch nichts beobachtbares Materielles bestimmt sind. Es ist leicht, zu sehen, dass wir damit die von Einstein aufgestellte kausale Forderung aufgeben, indem wir in eine ganz ähnliche Lage zurückkehren wie die, in welche die Annahme von Bewegungen in dem absoluten Raume oder in dem unterschiedslosen Weltäther als bestimmend für das Geschehen die Physik gebracht hatte.

Eddington, der augenscheinlich de Sitters Hypothese bevorzugt, erkennt denn auch ganz offen an, dass sie nicht der kausalen Relativitätsforderung, oder, wie er es nennt, dem philosophischen Prinzipe der Relativität gerecht wird. Aber er meint, dass es für den Physiker kein besonderes Interesse hat, dieses Prinzip zu behaupten, und dass wir uns damit begnügen müssen, »*to accept some mathematical expression as an axiomatic property which cannot be further analysed*« (»The Relativity Theory of Gravitation«, S. 91). Man könnte mit einigem Rechte sagen, dass er hierdurch ein mathematisches Ideal statt eines physikalischen Ideales als Leitstern für die physikalische Forschung anweist.

Sowohl Einsteins als de Sitters Sphärizitätshypothesen sind aus der Einstein'schen Relativitätstheorie als »Grenzbedingungen« entsprungen, aber keine von ihnen ist eine notwendige Folge von dieser. Von einem physikalischen Gesichtspunkte aus scheint mir die Hypothese von einer ins Unendliche fortgesetzten Systemordnung bei weitem vorzuziehen zu sein, nicht nur weil sie mit der kausalen Relativitätsforderung — von prinzipiell beobachtbaren physikalischen Ursachen — vollkommen harmoniert, sondern auch weil sie eine Fortsetzung von dem durch wirkliche Beobachtungen gewonnenen Weltbilde repräsentiert.

Gegen Einsteins Sphärizitätshypothese haben de Sitter und Eddington den Einwand erhoben, dass sie die Einführung von enormen Mengen von unbeobachtetem »Weltstoff« erfordert. Dieses wird überhaupt notwendig sein, wenn wir reale Massen für die Inertialsysteme als bestimmend ansehen wollen, denn die (auch nur mit Fernröhren) sichtbaren Sterne genügen gar nicht als felderzeugende Welt für unsere Gegend. Aber bei der »Systemhypothese« ist die Miteinbeziehung von viel grösseren Sternensystemen keine willkürliche oder künstliche Annahme (vergl. S. 53), und der genannte Einwand hat deshalb hier keine wirkliche Berechtigung. Es hängt teils von den Neutralfeldgesetzen, teils von der Verteilung der Sterne ab, zu welcher Grössenordnung der Systeme man gehen muss, um mit einer gewissen Genauigkeit den Vorgängen in einer gewissen Raumgegend gerecht zu werden, aber einen bestimmten endlichen Wert für die Masse und Ausdehnung der felderzeugenden Welt für eine gewisse Raumgegend gibt es nicht, und die einer bestimmten Genauigkeit entsprechende Masse und Raumgrösse der felderzeugenden Welt kann für verschiedene Raumgegenden verschieden sein. Einer bestimm-

ten »Weltmasse« oder einem bestimmten »Weltradius» wie in den Sphärizitätshypothesen begegnen wir also in keiner Weise in der Systemhypothese.

Was das Verhältnis der Systemhypothese zu der allgemeinen Einstein'schen Relativitätstheorie betrifft, leuchtet es nicht unmittelbar ein, dass sie nicht in einem solchen Weltbau Platz finden könnte. Die Systemhypothese steht wenigstens nicht in prinzipiellem Gegensatz zu den Grundgedanken der Einstein'schen Theorie. Eine andere Frage ist es, ob es möglich ist, die gesuchten und noch nicht festgestellten Grenzbedingungen dieser Theorie so zu sagen durch die Annahme der unendlichen Systemordnung zu ersetzen und auf diesem Weg die kausale Relativitätsforderung mit der vollständig durchgeführten Einstein'schen Relativitätstheorie zu vereinbaren. Die Beantwortung dieser Frage erfordert theoretische Untersuchungen, die ausserhalb des Rahmens dieser Abhandlung fallen. Es scheint mir aber, dass sie für die Auffassung der neueren Richtungen in der physikalischen Forschung eine grosse Bedeutung hat.

Für wertvolle Kritik und nützliche Ratschläge bei der Ausarbeitung dieser Abhandlung möchte ich gern den Herren Professoren K. Kroman und N. Bohr meinen besten Dank überbringen.

Charlottenlund, Dänemark
Mai 1919.

Færdig fra Trykkeriet den 18. Oktober 1919.

Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **II**, 12.

RECHERCHES

SUR LES

POLYNOMES DE STIRLING

PAR

NIELS NIELSEN

KØBENHAVN
HOVEDKOMMISSIONÆR: ANDR. FRED. HØST & SØN, KGL. HOF-BOGHANDEL
BIANCO LUNOS BOGTRYKKERI
1920

INTRODUCTION

JACOBUS STIRLING, dans son Traité [1] intéressant, mais trop peu connu, a étudié les deux classes de nombres positifs entiers que nous désignons, dans ce qui suit, par C_{n+1}^{p} et $\mathfrak{C}_{n+1}^{p}$, et l'étude de STIRLING est si profonde qu'il a calculé des petites tables et des C_{n+1}^{p} et des $\mathfrak{C}_{n+1}^{p}$. C'est pourquoi nous désignons comme nombres de STIRLING de première et de seconde espèce les C_{n+1}^{p} respectivement les $\mathfrak{C}_{n+1}^{p}$.

Les deux classes de positifs entiers, introduits par STIRLING, jouant un rôle assez fondamental et dans l'Analyse et dans la théorie des Nombres, ils ont été étudiés par beaucoup de géomètres; nous nous bornerons à citer ici EULER [2], LAPLACE [3], HERSCHEL [4], GRUNERT [5] et SCHLÖMILCH [6]. Dans nos jours, feu M. THIELE [7] a appliqué, du même point de vue que STIRLING, les C_{n+1}^{p} et les $\mathfrak{C}_{n+1}^{p}$, et à la demande de THIELE, l'éminent calculateur M. N.-P. BERTELSEN a donné une extension très considérable des petites tables de STIRLING. Le dernier article du présent Mémoire donne le commencement des tables de M. BERTELSEN.

[1] Methodus Differentialis: sive Tractatus de summatione et interpolatione serierum infinitarum. Londres 1730.

[2] Novi Commentarii Academiae Petropolitanae, t. 13. Institutiones calculi differentialis, p. 485—486; Pétrograde 1755.

[3] Histoire de l'Académie, année 1777, p. 99—122.

[4] Philosophical Transactions 1814 I, p. 440—468, 1816 I, p. 25—45.

[5] Mathematische Abhandlungen, p. 63—93. Altona 1822.

[6] Journal de Crelle, t. 44, p. 344—355, 1852.

[7] Interpolationsrechnung, p. 31—32. Leipsic 1909.

Les nombres de Stirling étant intimement liés aux nombres de Bernoulli, j'étudie depuis plus de trente ans, ces nombres compliqués, sur lesquels j'ai publié plusieurs notes, savoir:

Om Potenssummer af hele Tal.[1]

Recherches sur les polynomes et les nombres de Stirling.[2]

Note sur quelques applications analytiques des polynomes de Stirling.[3]

Sur les séries de fonctions de Stirling.[4]

Or, ayant mené à une terminaison relative mes recherches systématiques sur les nombres de Bernoulli[5], je me suis décidé à revenir aux nombres de Stirling, beaucoup plus difficiles encore.

En effet, l'étude systématique des nombres C_{n+1}^{p} et $\mathfrak{S}_{n+1}^{p}$, le rang p étant supposé fixe, est réduite à l'étude d'un seul polynome $\Phi_{2p}(x)$ du degré $2p$ par rapport à x, polynome qui se présente sous la forme

$$\Phi_{2p}(x) = (x+1)x(x-1)\ldots(x-p+1)\psi_{p-1}(x),$$

où $\psi_{p-1}(x)$ est un polynome entier du degré $p-1$ par rapport à x. C'est-à-dire que la détermination des nombres de Stirling du rang p renferme implicitement la détermination d'une factorielle de l'ordre p, inférieur à l'ordre des C_{n+1}^{p} et $\mathfrak{S}_{n+1}^{p}$.

Quant aux polynomes $\psi_p(x)$, que je désigne comme les polynomes de Stirling, soit

$$\psi_p(x) = \alpha_{p,0}x^p + \alpha_{p,1}x^{p-1} + \ldots + \alpha_{p,r}x^{p-r} + \ldots + \alpha_{p,p},$$

[1] Nyt Tidsskrift for Mathematik, t. 4 B, p. 1—10; 1893.

[2] Annali di Matematica, 3e série, t. 10, p. 287—318; 1904.

[3] Id. p. 319—325; 1904.

[4] Annali di Matematica, 3e série, t. 12, p. 101—112; 1905.

[5] La guerre a retardé la publication de mon livre: *Traité élémentaire des Nombres de Bernoulli.* Dans le Mémoire présent je cite plusieurs fois ce Traité, en indiquant les numéros des articles dont il s'agit.

et soit r un nombre fixe, il existe un polynome $\beta_{2r}(x)$ du degré $2r$ par rapport à x, tel que

$$\alpha_{p,r} = \frac{\beta_{2r}(p)}{(p+1)!\,2^{p+1}}, \qquad 0 \leq r < p,$$

et ce polynome $\beta_{2r}(x)$ se présente sous la forme

$$\beta_{2r}(x) = x(x-1)\dots(x-r+1)\,\sigma_r(x),$$

où $\sigma_r(x)$ est un nouveau polynome entier du degré r par rapport à x. C'est-à-dire que la détermination du coefficient $\beta_{p,r}$ exige aussi la détermination d'une factorielle, étant de l'ordre r inférieur à p.

Ces remarques montrent clairement que la nature analytique des nombres de Stirling est très compliquée.

De plus, j'ai donné une théorie élémentaire des nombres de Bernoulli, en les débarrassant entièrement des éléments transcendentaux, mais je n'ai pas réussi à résoudre ce même problème relativement aux nombres de Stirling. C'est pourquoi j'ai été obligé d'appliquer, dans le présent Mémoire, et une méthode élémentaire et une méthode transcendante.

Quant au Mémoire que j'ai l'honneur de présenter ici à notre Académie, il contient tous les résultats donnés dans mes publications susdites, souvent dans une forme généralisée, et beaucoup d'autres. Or, je ne mentionne pas les séries de factorielles de Stirling, données dans mon Traité de la fonction gamma[1], et c'est la même chose pour la belle généralisation de ces formules, savoir

$$\frac{\Gamma(x)\,\Gamma(y)}{\Gamma(y-x+1)\,(y-x)^x} = \frac{\Gamma(x)}{y} + \sum_{s=1}^{s=\alpha} \frac{\Gamma(x+s+1)\,\psi_{s-1}(x+s-1)}{y(y+1)\dots(y+s)};$$

[1] Handbuch der Theorie der Gammafunktion, pp. 259, 285—288; Leipsic 1906.

où les $\psi_{s-1}(x+s-1)$ sont les polynomes de Stirling et où il faut supposer

$$\Re(y) > \Re(x) > -1.$$

Cette formule est indiquée par M. Rutgers[1], mais sa démonstration, une intégration terme à terme d'une série infinie, n'est malheureusement pas rigoureuse, et je n'ai pas réussi à la compléter. De plus, il ne me semble pas sûr que les conditions indiquées par M. Rutgers soient exactes.

En appliquant, dans les articles VIII, IX, X du présent Mémoire, des séries de Lagrange, dans lesquelles les nombres de Stirling jouent un rôle fondamental, je saisis l'occasion d'attirer l'attention des géomètres sur une fausseté littéraire qui rattache le nom de Bürmann aux séries de la forme

$$\text{(a)} \quad f(x) = A_0 + A_1\varphi(x) + A_2\big(\varphi(x)\big)^2 + \ldots + A_n\big(\varphi(x)\big)^n + \ldots,$$

appliquables aux environs de $x = a$, zéro simple de la fonction analytique $\varphi(x)$, et dont les coefficients sont à déterminer par les expressions

$$\text{(b)} \qquad \begin{cases} A_0 = f(a) \\ A_n = \dfrac{1}{n!} D_x^{n-1}\left[\left(\dfrac{x-a}{\varphi(x)}\right)^n f'(x)\right]_{x=a}. \end{cases}$$

La source de cette fausseté littéraire est peut-être à chercher dans une citation trompeuse de Lacroix[2], citation qui fait croire que Bürmann ait publié un mémoire dans le tome II, p. 15 des Mémoires de l'Institut.

Or, le tome susdit (de l'an IV) ne contient qu'une recension, faite par Lagrange et soussignée également par

[1] Nieuw archief voor Wiskunde, 3e série, t. 8, p. 106; 1907.

[2] Traité du calcul différentiel et du calcul intégral, t. III, p. XXI (Table), 2e édition, Paris 1819.

LEGENDRE, de deux mémoires manuscrits que BÜRMANN a envoyés à l'Institut. Et, dans la page citée par LACROIX, on lit textuellement:

»Cette formule (savoir une formule équivalente à (a), suppléée par les expressions (b) des coefficients) étant très générale et susceptible d'un grand nombre d'applications utiles dans la théorie des fonctions, nous avons cherché si elle ne s'étoit pas présentée déjà aux regards des analystes: nous avons bientôt reconnu qu'elle ne différoit pas essentiellement du théorème donné par Lagrange dans les *Mémoires de Berlin*, année 1769 [1768?]. Nous avons trouvé également qu'elle pouvoit se déduire très aisément d'un théorème que le citoyen Laplace a donné sans démonstration dans les *Mémoires de l'Académie des Sciences* année 1777. Il résulte de là que la formule de M. Bürmann ne peut être regardée comme une découverte nouvelle en analyse; mais on ne sauroit refuser à l'auteur la justice de reconnoître qu'il est parvenu à ce théorème par ses propres recherches, et qu'il en a donné une démonstration nouvelle et ingénieuse.«

Quant à la citation inexacte de LACROIX, je remarque expressément que cet auteur désigne [1] la formule en question comme théorème de LAGRANGE.

Examinons maintenant la littérature plus ancienne que le Traité susdit de LACROIX.

KLÜGEL, dans son Dictionnaire Mathématique [2] ne mentionne pas BÜRMANN, mais indique correctement que LAGRANGE [3] a démontré le théorème général:

[1] Traité du calcul différentiel et du calcul intégral, t. I, p. IV (Table), 2e édition. Paris 1810.

[2] Mathematisches Handwörterbuch, t. I, p. 625—636; Leipsic 1802.

[3] Mémoires de l'Académie de Berlin, année 1768, p. 275; voir aussi année 1769, p. 207.

Soit x une racine quelconque de l'équation

$$(c) \qquad a = x - \psi(x),$$

où $\psi(x)$ est une fonction »quelconque« de x, et soit $f(x)$ une autre fonction »quelconque« de x, il résulte le développement

$$(d) \qquad f(x) = f(a) + \sum_{n=1}^{n=\infty} \frac{1}{n!} D_x^{n-1} \left[(\psi(x))^n f'(x) \right]_{x=a}.$$

De plus, Klügel [1] remarque, qu'en substituant, dans (c), $y\psi(x)$ au lieu de x, savoir en étudiant cette autre équation

$$(e) \qquad a = x - y\psi(x),$$

on aura plus généralement

$$(f) \qquad f(x) = f(a) + \sum_{n=1}^{n=\infty} \frac{y^n}{n!} D_x^{n-1} \left[(\psi(x))^n f'(x) \right]_{x=a},$$

généralisation immédiate qui était donnée déjà par Lagrange. [2]

Soit maintenant $y = \varphi(x)$, ce qui donnera, en vertu de (e),

$$\psi(x) = \frac{x-a}{\varphi(x)},$$

on voit que la formule (a) est une conséquence immédiate de (f), la détermination (b) des coefficients A_n y comprise.

Remarquons encore que J.-W. Pfaff [3] en reproduisant la dernière démonstration de Lagrange, ne mentionne pas Bürmann et c'est la même chose pour Grunert, dans le second volume du Dictionnaire Mathématique [4], et pour

[1] Loc. cit. p. 625.

[2] Théorie des fonctions analytiques, p. 101—105; Paris, an V (1797).

[3] Sammlung combinatorisch-analytischer Abhandlungen, herausgegeben von Carl Friedrich Hindenburg. Zweyte Sammlung, p. 195—229: Leipsic 1800.

[4] Mathematisches Handwörterbuch, Supplement Bd. I, pp. 55, 80; Leipsic 1833.

Puiseux[1], dans son Mémoire fondamental sur les fonctions algébriques.

C. Ramus[2], l'érudit professeur danois, indique, sans donner des citations, mais évidemment inspiré par Lacroix, que les formules (b) sont dues à Bürmann, ce qui est faux, mais Ramus dit expressément, que la série (a) n'est autre chose qu'une transformation immédiate de celle de Lagrange.

La désignation: *série de Bürmann* est appliqué eexplicitement par Schlömilch[3] qui, à cette occasion, fait la remarque suivante:

»Die Formeln 62), 63) und 64) [savoir les formules (a) et (b)] wurden 1796 von Bürmann (in Mannheim) auf anderem Wege entwickelt (Mémoire de l'Institut, tome II, pag. 14) jedoch, dem damaligen Standpunkte der Wissenschaft gemäss, ohne Angabe der Gültigkeitsbedingungen für die Gleichung 62) [savoir la formule (a)]. Neuerdings ist Puiseux (Liouville's Journal Bd. 15) etc.«

On voit que cette remarque du savant géomètre saxon est très étourdie, et il est évident que Schlömilch n'a jamais vu la page 14 du tome II des *Mémoires de l'Institut!*

Quant à l'inexactitude (certaine, mais pas mise en pleine lumière parce que le mémoire n'est pas publié) des conditions de Bürmann, on est tenté de désigner comme des fantaisies sa propre démonstration. Et, sous ce point de vue, c'est très curieux, ce me semble, que Schlömilch mentionne la démonstration rigoureuse et moderne de Puiseux. Or cette démonstration étant désignée comme récente, il est évident que la remarque de Schlömilch est écrite peu de temps après 1850.

[1] Journal de Liouville, t. 15, p. 365—480, 1850 (voir p. 380—383).

[2] Algebra og Functionslære, p. 108—112; Copenhague 1840.

[3] Compendium der höheren Analysis, t. II, p. 102 (3e édition); Brunswick 1879.

Curieusement, Klügel remarque que Condorcet, dans l'Encyclopédie Méthodique (article séries), attribue à d'Alembert la série de Lagrange.

Quant à Bürmann, désigné, dans la recension susdite, comme professeur de commerce à Manheim, je sais seulement qu'il a publié, dans l'Archiv für reine und angewandte Mathematik, t. II, un mémoire intitulé: *Berechnung des Kreises.*

Copenhague, le 26 juin 1919.

Niels Nielsen.

PREMIÈRE PARTIE

Les nombres de Stirling.

I. — Définitions et propriétés fondamentales.

Soit n un nombre positif, les factorielles des ordres $\pm n$ sont à définir comme suit

$$(1)\quad \begin{cases} \omega_n(x) = x(x+1)(x+2)\dots(x+n-1) \\ \omega_{-n}(x) = \dfrac{1}{x(x+1)(x+2)\dots(x+n-1)}; \end{cases}$$

de plus, nous posons

$$(1\text{ bis})\qquad \omega_0(x) = 1.$$

Étudions tout d'abord la factorielle d'un ordre positif, puis posons

$$(2)\quad \begin{cases} \omega_{n+1}(x) = \\ C_{n+1}^0 x^{n+1} + C_{n+1}^1 x^n + \dots + C_{n+1}^p x^{n-p+1} + \dots + C_{n+1}^n x, \end{cases}$$

les positifs entiers C_{n+1}^p sont désignés comme les coefficients de factorielle, ou les nombres de STIRLING de première espèce, de l'ordre $n+1$. Nous aurons évidemment

$$(3)\quad \begin{cases} C_n^0 = 1, \qquad n \geqq 0 \\ C_{n+1}^n = n!, \end{cases}$$

tandis que le coefficient général C_{n+1}^p est la somme des produits formés de p facteurs inégaux pris parmi les nombres

$$1, 2, 3, \dots, n.$$

Quant aux factorielles d'un ordre négatif, je dis que nous aurons un développement de la forme

$$\omega_{-n-1}(x) = \sum_{s=0}^{s=\infty} \frac{(-1)^s \mathfrak{S}_{n+1}^{s}}{x^{n+s+1}\,\lfloor s \rfloor}, \qquad |x| > n, \tag{4}$$

où les coefficients $\mathfrak{S}_{n+1}^{r}$, les coefficients de factorielle, ou les nombres de STIRLING de seconde espèce, de l'ordre $-n-1$, sont déterminés par les expressions

$$\mathfrak{S}_{n+1}^{r} = \frac{1}{n!} \cdot \sum_{s=0}^{s=n} (-1)^s \binom{n}{s} (n-s)^{n+r}. \tag{5}$$

En effet, prenons pour point de départ l'identité evidente

$$\frac{1}{x(x+1)\dots(x+n)} = \frac{1}{n!} \cdot \sum_{s=0}^{s=n} \frac{(-1)^s}{x+s} \binom{n}{s}, \tag{6}$$

puis appliquons la série géométrique

$$\frac{1}{x+s} = \frac{1}{x} - \frac{s}{x^2} + \frac{s^2}{x^3} - \dots + \frac{(-1)^m s^m}{x^{m+1}} + \dots, \tag{6 bis}$$

valable pour $1 \leq s \leq n$, pourvu que $|x| > n$, le coefficient de la puissance

$$\frac{(-1)^m}{x^{m+1}},$$

qui figure dans l'expression, obtenue à l'aide de (6), deviendra

$$\frac{1}{n!} \cdot \sum_{s=0}^{s=n} (-1)^s \binom{n}{s} (n-s)^m. \tag{7}$$

Cela posé, les valeurs limites évidentes

$$\lim_{x=\infty} \left(x^m \omega_{-n-1}(x) \right) = 0, \qquad 1 \leq m < n, \tag{7 bis}$$

montrent clairement que l'expression (7) s'évanouira pour $1 \leq m < n$, ce qui donnera immédiatement la formule (4). Remarquons que la valeur limite (7 bis) a, pour $m = n+1$, la valeur 1, nous aurons de plus

$$\mathfrak{S}_n^0 = 1, \qquad n \geq 0. \tag{8}$$

Curieusement, M. Unferdinger[1] désigne comme théorème de Legendre cette dernière formule; or, ce résultat numérique est certainement plus ancien.

Soit ensuite α un nombre complexe quelconque, nous posons plus généralement

$$(9)\quad \begin{cases} \omega_{n+1}(x+\alpha) = \\ = C^0_{n+1}(\alpha)x^{n+1} + C^1_{n+1}(\alpha)x^n + \ldots + C^n_{n+1}(\alpha)x + C^{n+1}_{n+1}(\alpha), \end{cases}$$

$$(9\text{ bis})\qquad \omega_{-n-1}(x+\alpha) = \sum_{s=0}^{s=\infty} \frac{(-1)^s \mathfrak{C}^s_{n+1}(\alpha)}{x^{n+s+1}},$$

où il faut supposer, dans cette dernière formule,

$$|x| > |\alpha + s|, \qquad 0 \leq s \leq n,$$

il résulte évidemment

$$(10)\quad \begin{cases} C^r_{n+1}(0) = C^r_{n+1}, & 0 \leq r \leq n, \\ \mathfrak{C}^r_{n+1}(0) = \mathfrak{C}^r_{n+1}, & r > 0. \end{cases}$$

Quant au coefficient général $C^p_{n+1}(\alpha)$, il est la somme des produits de p facteurs inégaux pris parmi les nombres

$$\alpha,\ \alpha+1,\ \alpha+2,\ \ldots,\ \alpha+n;$$

ce qui donnera particulièrement

$$(11)\qquad C^{n+1}_{n+1}(\alpha) = \omega_{n+1}(\alpha).$$

Appliquons ensuite les séries géométriques

$$\frac{1}{x+\alpha+s} = \frac{1}{x} - \frac{\alpha+s}{x^2} + \frac{(\alpha+s)^2}{x^3} - \ldots, \qquad |x| > |\alpha+s|,$$

nous aurons de même

$$(12)\quad \mathfrak{C}^p_{n+1}(\alpha) = \frac{1}{n!} \cdot \sum_{s=0}^{s=n} (-1)^s \binom{n}{s} (\alpha+n-s)^{n+p},$$

[1] Wiener Sitzungsberichte t. 67 II, p. 365; 1873.

de sorte que $\mathfrak{S}^p_{n-1}(\alpha)$ n'est, abstraction faite d'un simple facteur, autre chose que la $n^{\text{ième}}$ différence de $(\alpha+n)^{n-p}$, savoir

$$\mathfrak{S}^p_{n+1}(\alpha) = \frac{1}{n!} \cdot \Delta^n (\alpha+n)^{n+p}. \tag{12 bis}$$

Cela posé, ordonnons, d'après des puissances descendantes de α, le second membre de (12), la formule binomiale donnera

$$\mathfrak{S}^r_{n+1}(\alpha) = \sum_{s=0}^{s=r} \binom{n+r}{r-s} \mathfrak{S}^s_{n+1}\, \alpha^{r-s}. \tag{13}$$

Quant aux coefficients $C^r_{n-1}(\alpha)$, appliquons l'identité

$$\omega_{n+1}(x+\alpha) = \sum_{s=0}^{s=n} C^s_{n-1}(x+\alpha)^{n-s+1},$$

tirée directement de la définition (2), la même méthode donnera

$$C^r_{n+1}(\alpha) = \sum_{s=0}^{s=r} \binom{n-s+1}{r-s} C^s_{n+1}\, \alpha^{r-s}. \tag{13 bis}$$

On voit que la formule (11) détermine les coefficients $\mathfrak{S}^r_{n+1}(\alpha)$ par des expressions explicites, tandis que les $C^r_{n+1}(\alpha)$ ne sont définis que par des méthodes combinatoires. Or, il faut ajouter que les expressions explicites susdites sont peu commodes pour un calcul direct des $\mathfrak{S}^r_{n-1}(\alpha)$. C'est pourquoi nous avons à chercher des formules récursives qui nous permettent de calculer successivement et les $C^r_{n+1}(\alpha)$ et les $\mathfrak{S}^r_{n-1}(\alpha)$.

A cet effet, appliquons les identités évidentes

$$(x+\alpha+n)\,\omega_n(x+\alpha) = \omega_{n+1}(x+\alpha)$$
$$(x+\alpha+n)\,\omega_{-n-1}(x+\alpha) = \omega_{-n}(x+\alpha),$$

puis cherchons, aux deux membres de ces formules, le coefficient de la puissance x^{n-r-1} respectivement x^{n-r}, il résulte, pour les $C_{n+1}^r(\alpha)$,

$$(14)\quad \begin{cases} C_{n+1}^r(\alpha) = C_n^r(\alpha) + (\alpha+n)\, C_n^{r-1}(\alpha), & 1 \leqq r \leqq n \\ C_{n+1}^0(\alpha) = C_n^0(\alpha) = 1 \\ C_{n+1}^{n+1}(\alpha) = (\alpha+n)\, C_n^n(\alpha) = \omega_{n+1}(\alpha), \end{cases}$$

et pour les $\mathfrak{S}_{n+1}^r(\alpha)$

$$(14\text{ bis})\quad \begin{cases} \mathfrak{S}_{n+1}^0(\alpha) = \mathfrak{S}_n^0(\alpha) = 1 \\ \mathfrak{S}_n^r(\alpha) = \mathfrak{S}_{n+1}^r(\alpha) - (\alpha+n)\, \mathfrak{S}_{n+1}^{r-1}(\alpha), & r \geqq 1, \end{cases}$$

d'où, en supposant $\alpha = 0$,

$$(15)\quad \begin{cases} C_{n+1}^r = C_n^r + n\, C_n^{r-1}, & 1 \leqq r \leqq n-1 \\ C_{n+1}^n = n C_n^{n-1} = n! \\ C_{n+1}^0 = C_n^0 = 1, \quad n > 0; \quad C_1^r = 0, \quad r \geqq 1, \end{cases}$$

et il est évident que la première de ces formules, suppléée par la valeur initiale, indiquée dans la dernière, détermine successivement tous les nombres de Stirling de première espèce.

Quant aux nombres $\mathfrak{S}_{n+1}^r$, il résulte de même

$$(15\text{ bis})\quad \begin{cases} \mathfrak{S}_n^r = \mathfrak{S}_{n+1}^r - n\, \mathfrak{S}_{n+1}^{r-1}, & r \geqq 1 \\ \mathfrak{S}_{n+1}^0 = \mathfrak{S}_n^0 = 1, \quad n \geqq 0; \quad \mathfrak{S}_1^r = 0, \quad r \geqq 1, \end{cases}$$

et il est évident que ces formules déterminent successivement tous les nombres de Stirling de seconde espèce, ce qui donnera la proposition suivante:

I. Les nombres de Stirling de seconde espèce, savoir les $\mathfrak{S}_{n+1}^r$, sont pour $n \geqq 1$, des positifs entiers.

II.—Formules récursives générales.

La définition même de la faculté d'un ordre quelconque nous permet de développer une suite de formules récursives beaucoup plus générales que les précédentes.

En premier lieu, prenons pour point de départ les identités évidentes

$$\frac{1}{x+\alpha+n}\cdot\omega_{n+1}(x+\alpha) = \omega_n(x+\alpha)$$

$$\frac{1}{x+\alpha+n}\cdot\omega_{-n}(x+\alpha) = \omega_{-n-1}(x+\alpha),$$

puis appliquons la série géométrique

$$\frac{1}{x+\alpha+n} = \frac{1}{x} - \frac{\alpha+n}{x^2} + \frac{(\alpha+n)^2}{x^3} - \dots, \quad |x| > |\alpha+n|,$$

nous aurons respectivement

$$(1) \qquad \sum_{s=0}^{s=r}(-1)^s(\alpha+n)^s C_{n+1}^{r-s}(\alpha) = C_n^r(\alpha), \quad 0 \leqq r \leqq n,$$

$$(1\text{ bis}) \qquad \sum_{s=0}^{s=r}(\alpha+n)^s \mathfrak{S}_n^{r-s}(\alpha) = \mathfrak{S}_{n+1}^r(\alpha), \qquad r \geqq 0.$$

On voit que la dernière de ces deux formules donnera, pour $\alpha = 0$, une nouvelle démonstration de la proposition I de l'article précédent concernant la nature des nombres de STIRLING de seconde espèce.

Quant à la formule (1), elle peut être suppléée par cette autre

$$(2) \qquad \sum_{s=0}^{s=n+1}(-1)^s(\alpha+n)^s C_{n+1}^{n-s+1}(\alpha) = 0,$$

obtenue en cherchant, dans le premier des produits susdits, le coefficient d'une puissance négative de x.

D'autres formules récursives plus importantes sont des conséquences immédiates des deux identités évidentes

$$\omega_{n+p+1}(x)\,\omega_{-n-1}(x) = \omega_p(x+n+1)$$

$$\omega_{-n-p-1}(x)\,\omega_{n+1}(x) = \omega_{-p}(x+n+1),$$

ou, ce qui est la même chose,

$$\left(\sum_{s=0}^{s=n+p} C^s_{n+p+1}\,x^{n+p-s+1}\right)\cdot\left(\sum_{s=0}^{s=\infty}\frac{(-1)^s\,\mathfrak{S}^s_{n+1}}{x^{n+s+1}}\right) = \sum_{s=0}^{s=p-1} C^s_p\,(n+1)\,x^{p-s}$$

$$\left(\sum_{s=0}^{s=\infty}\frac{(-1)^s\,\mathfrak{S}^s_{n+p+1}}{x^{n+p+s+1}}\right)\cdot\left(\sum_{s=0}^{s=n} C^s_{n+1}\,x^{n-s+1}\right) = \sum_{s=0}^{s=\infty}\frac{(-1)^s\,\mathfrak{S}^s_p\,(n+1)}{x^{p+s}}.$$

Cherchons maintenant, aux deux membres de ces formules, le coefficient de la puissance x^{p-r} respectivement x^{-p-r}, les seconds membres donnent $C^r_p(n+1)$, et $(-1)^r\mathfrak{S}^r_p(n+1)$, d'où, en vertu des formules, obtenues des (13) et (13 bis) de l'article précédent, en y posant $n = p-1$, $\alpha = n+1$,

$$(3)\qquad \sum_{s=0}^{s=r}(-1)^s\,C^{r-s}_{n+p+1}\,\mathfrak{S}^s_{n+1} = \sum_{s=0}^{s=r}\binom{p-s}{r-s}(n+1)^{r-s}\,C^s_p$$

$$(3\text{ bis})\qquad \sum_{s=0}^{s=r}(-1)^s\,\mathfrak{S}^{r-s}_{n+p+1}\,C^s_{n+1} = \sum_{s=0}^{s=r}\binom{p+r-1}{r-s}(n+1)^{r-s}\,\mathfrak{S}^s_p.$$

Dans ces deux formules il faut, pour des valeurs plus grandes de r, supprimer les termes contenant des coefficients C^m_q, dans lesquels $m \geq q$. Soit particulièrement $n = 0$, les premiers membres des deux formules en question se réduisent à leur premier terme, ce qui donnera respectivement

$$(4)\qquad C^r_{p+1} = \sum_{s=0}^{s=r}\binom{p-s}{r-s}C^s_p$$

$$(4\text{ bis})\qquad \mathfrak{S}^r_{p+1} = \sum_{s=0}^{s=r}\binom{p+r-1}{r-s}\mathfrak{S}^s_p.$$

Une autre formule essentielle peut être déduite de (3) en y posant $r > p$, de sorte que le second membre de cette formule s'évanouira, d'où, en remplaçant $n + p$ par m,

$$(5) \qquad \sum_{s=0}^{s=r} (-1)^s C_{m+1}^{r-s} \mathfrak{C}_{n+1}^s = 0, \qquad r > m-n > 0,$$

tandis que l'hypothèse $r = p$, savoir $r = m-n$, donnera

$$(5\text{ bis}) \qquad \sum_{s=0}^{s=r} (-1)^s C_{n+r+1}^{r-s} \mathfrak{C}_{n+1}^s = \omega_r(n+1) = \frac{(n+r)!}{n!}.$$

Revenons maintenant aux développements (2) et (4) de l'article précédent, puis posons $x-1$ au lieu de x, nous aurons, en divisant, respectivement multipliant par $x-1$,

$$\omega_n(x) = \sum_{s=0}^{s=n} C_{n+1}^s (x-1)^{n-s}$$

$$\omega_{-n}(x) = \sum_{s=0}^{s=\infty} \frac{(-1)^s \mathfrak{C}_{n+1}^s}{(x-1)^{n+s}},$$

d'où, en cherchant, aux deux membres de ces formules, le coefficient de la puissance x^{n-r} respectivement x^{-n-r},

$$(6) \qquad C_n^r = \sum_{s=0}^{s=r} (-1)^s \binom{n-r+s}{s} C_{n+1}^{r-s}$$

$$(6\text{ bis}) \qquad \mathfrak{C}_n^r = \sum_{s=0}^{s=r} (-1)^s \binom{n+r-s}{s} \mathfrak{C}_{n+1}^{r-s},$$

formules qui représentent les inversions des formules (4).

Posons ensuite, dans (6) et (6 bis), $r+1$ au lieu de r, puis éliminons, en vertu des formules récursives (15) et (15 bis) de l'article précédent, C_{n+1}^{r-1} respectivement $\mathfrak{C}_{n+1}^{r+1}$, il résulte

$$(7) \qquad n C_n^r = \sum_{s=0}^{s=r} (-1)^s \binom{n+r+s}{s+1} C_{n+1}^{r+s}$$

$$(7\text{ bis}) \qquad n \mathfrak{C}_n^r = \sum_{s=0}^{s=r} (-1)^s \binom{n+r-s}{s+1} \mathfrak{C}_{n+1}^{r-s},$$

tandis que la même méthode donnera, en vertu de (4),

$$(8) \qquad n C_n^r = \sum_{s=0}^{s=r} \binom{n-s}{r-s+1} C_n^s$$

$$(8\text{ bis}) \qquad n \mathfrak{C}_{n+1}^r = \sum_{s=0}^{s=r} \binom{n+r}{r-s+1} \mathfrak{C}_n^s,$$

formules qui représentent les inversions des formules (7).

Dans l'article XII nous avons à donner des applications intéressantes des développements numériques que nous venons d'établir ici.

III.—Applications élémentaires.

Les formules numériques, développées dans l'article précédent, conduisent immédiatement à des applications intéressantes des nombres de Stirling.

En premier lieu, prenons pour point de départ la formule (5) de l'article susdit, il est très facile de démontrer le théorème:

I. Cherchons, des équations linéaires

$$(1) \quad \begin{cases} y_1 = C_1^0 x_1, \\ y_2 = C_2^0 x_2 + C_2^1 x_1, \\ y_3 = C_3^0 x_3 + C_3^1 x_2 + C_3^2 x_1, \\ \dots\dots\dots\dots\dots\dots \\ y_n = C_n^0 x_n + C_n^1 x_{n-1} + C_n^2 x_{n-2} + \dots + C_n^{n-1} x_1, \end{cases}$$

les quantités $x_1 x_2 x_3 \ldots x_n$, il résulte

$$(1\text{ bis})\quad \begin{cases} x_1 = \mathfrak{S}_2^0\, y_1 \\ x_2 = \mathfrak{S}_3^0\, y_2 - \mathfrak{S}_2^1\, y_1 \\ x_3 = \mathfrak{S}_4^0\, y_3 - \mathfrak{S}_3^1\, y_2 + \mathfrak{S}_2^2\, y_1 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ x_n = \mathfrak{S}_{n+1}^0\, y_n - \mathfrak{S}_n^1\, y_{n-1} + \ldots + (-1)^{n-1}\mathfrak{S}_2^{n-1}\, y_1, \end{cases}$$

et inversement.

En effet, introduisons, dans la formule générale (1), les expressions indiquées dans (1 bis), la formule susdite de l'article précédent donnera une identité. De même, introduisons, dans la formule générale (1 bis) les expressions indiquées dans (1), nous aurons aussi une identité.

Quant aux nombres $\mathfrak{S}_{n+1}^r$, il est facile de démontrer le théorème:

II. Développons, d'après la formule polynomiale, l'expression

$$(a_1 + a_2 + a_3 + \ldots + a_p)^n,$$

où n et p désignent des positifs entiers, puis désignons par $\Omega_{p,n}$ la somme des coefficients polynomiaux qui correspondent aux termes divisibles par le produit $a_1 a_2 \ldots a_p$, nous aurons généralement

$$(2)\qquad \Omega_{p,n} = p!\,\mathfrak{S}_{p+1}^{n-p}.$$

Soit particulièrement $p = 2$, on aura directement

$$\Omega_{2,n} = 2^n - 2 = 2!\,\mathfrak{S}_3^{n-2};$$

appliquons ensuite l'identité évidente

$$(a_1 + a_2 + a_3 + \ldots + a_p)^n = \sum_{s=0}^{s=n} \binom{n}{s} a_1^s (a_2 + a_3 + \ldots + a_p)^{n-s},$$

il résulte, pour les $\Omega_{p,n}$, la formule récursive

$$\Omega_{p,n} = \sum_{s=1}^{s=n-p+1} \binom{n}{s} \Omega_{p-1,n-s},$$

et la conclusion de n à $n+1$ est une conséquence immédiate de la formule (8 bis) de l'article précédent.

Dans la formule (12 bis) de l'article I nous avons développé la différence $\Delta^n (x+n)^p$ d'après des puissances descendantes de x; le calcul aux différences finies donnera aussi facilement cette autre formule

$$(3) \qquad x^n = \sum_{s=0}^{s=n-1} (-1)^s \mathfrak{S}^s_{n-s+1}\, \omega_{n-s}(x)$$

qui peut être considérée comme l'inversion de la définition de $\omega_n(x)$, savoir la formule (2) de l'article I. Dans l'article IX nous avons à déduire, d'un autre point de vue, le développement (3).

IV.—Introduction d'une variable continue.

Remarquons que l'équation algébrique du $(n+1)^{\text{ième}}$ degré

$$\omega_{n+1}(x) = x^{n+1} + C^1_{n+1} x^n + C^2_{n+1} x^{n-1} + \ldots + C^n_{n+1} x = 0$$

a les $n+1$ racines

$$0,\ -1,\ -2,\ -3,\ \ldots,\ -n,$$

puis posons pour abréger

$$(1) \qquad S_p(n) = 1^p + 2^p + 3^p + \ldots + n^p,$$

il résulte, en vertu des formules de Newton,

$$(2) \qquad \sum_{r=0}^{r=p-1} (-1)^r C^r_{n+1} S_{p-r}(n) + (-1)^p p\, C^p_{n+1} = 0, \qquad 1 \leq p \leq n,$$

formules qui nous permettent d'introduire, dans l'étude des nombres de Stirling, une variable continue au lieu du positif entier n.

A cet effet, posons

$$(3)\quad \begin{cases} \Phi_1(x) = x \\ \Phi_2(x) = \dfrac{x^2}{2} + \dfrac{x}{2} \\ \Phi_p(x) = \dfrac{x^p}{p} + \dfrac{x^{p-1}}{2} + \sum\limits_{s=1}^{\leqq \frac{p-1}{2}} \dfrac{(-1)^{s-1} B_s}{p} \binom{p}{2s} x^{p-2s}, \end{cases}$$

où les B_m désignent les nombres de Bernoulli, tandis que x est une variable complexe quelconque, Jacques Bernoulli[1] a indiqué la formule générale

$$(4)\qquad S_{p-1}(n) = \Phi_p(n), \qquad p \geqq 2,$$

où n désigne un positif entier quelconque.

Cela posé, étudions le système d'équations linéaires

$$(5)\quad \sum_{s=0}^{s=p-1} (-1)^s \Phi_{p-s-1}(x)\, \Psi_{2s}(x) + (-1)^p p\, \Psi_{2p}(x) = 0,$$

où nous supposons

$$(5\text{ bis})\qquad \Psi_0(x) = 1,$$

il est évident que ces équations deviennent identiques aux formules numériques (2), si nous posons $x = n$, où n est un positif entier, égal à p au moins. Or, je dis que $\Psi_{2p}(x)$, déterminé par les équations (5), est un polynome entier du degré $2p$ par rapport à x.

En effet, supposons vraie cette propriété de $\Psi_{2r}(x)$ pour $1 \leqq r \leqq p-1$, il est évident que le terme

$$(-1)^{p-1} \Phi_2(x)\, \Psi_{2p-2}(x)$$

est un polynome entier du degré $2p$ par rapport à x,

[1] Ars conjectandi, p. 97; Bâle 1713.

tandis que tous les autres termes figurant sous le signe sommatoire au premier membre de la formule susdite sont d'un degré inférieur à $2p$.

De plus, nous aurons, en vertu de (5),

$$\Psi_{2p}(0) = 0, \quad p > 1. \tag{6}$$

Cela posé, il est facile de démontrer une suite de théorèmes fondamentaux concernant les polynomes $\Psi_{2p}(x)$, savoir:

I. La suite des polynomes $\Psi_{2p}(x)$ est une généralisation des nombres de Stirling et de première et de seconde espèce, parce que nous aurons, en désignant par n un positif entier,

$$\Psi_{2p}(n) = C^p_{n+1}, \quad n \geq p, \; p \geq 0, \tag{7}$$

$$\Psi_{2p}(-n-1) = \mathfrak{S}^p_{n+1}, \quad p \geq 0. \tag{7 bis}$$

La première de ces deux formules est évidente, parce que les deux systèmes d'équations (5) et (2) deviennent identiques pour $x = n \geq p \geq 1$. Quant à la seconde des formules en question, je dis que les polynomes $\Psi_{2p}(x)$ satisfont à l'équation fonctionnelle:

$$\Psi_{2p}(x+1) = \Psi_{2p}(x) + (x+1)\Psi_{2p-2}(x), \quad p > 1. \tag{8}$$

En effet, soit x égal au positif entier $n \geq p$, l'équation (8) n'est, en vertu de (7), autre chose que la formule récursive (15) de l'article I; c'est-à-dire que l'équation algébrique (8), du degré $2p$ au plus, admet une infinité de racines inégales.

Posons ensuite, dans l'identité (8), $-x-1$ au lieu de x, il résulte

$$\Psi_{2p}(-x) = \Psi_{2p}(-x-1) - x\Psi_{2p-2}(-x-1),$$

ce qui donnera immédiatement, en vertu de (15 bis) de

l'article I, la formule (7 bis), formule qui est évidente pour $p = 0$.

II. Les équations fonctionnelles (8), suppléées par la valeur initiale $\Psi_0(x) = 1$ et par la condition $\Psi_{2p}(0) = 0$, déterminent parfaitement la suite des polynomes

$$\Psi_0(x),\ \Psi_2(x),\ \Psi_4(x),\ \ldots,\ \Psi_{2n}(x),\ \ldots,$$

définis par les formules (5).

En effet, supposons donné le polynome $\Psi_{2p-2}(x)$, la formule (8) est une équation linéaire aux différences finies, et cette équation du premier ordre détermine le polynome $\Psi_{2p}(x)$, abstraction faite d'une constante additive, dont la valeur est déterminée par la condition $\Psi_{2p}(0) = 0$.

Dans ce qui suit nous avons à appliquer indistinctement les deux définitions que nous venons de donner pour les polynomes $\Psi_{2p}(x)$.

III. Les coefficients $a_{n,r}$ du polynome

$$(9)\quad \begin{cases} \Psi_{2n}(x) = a_{n,0}\,x^{2n} + a_{n,1}\,x^{2n-1} + \ldots + a_{n,p}\,x^{2n-p} + \ldots \\ \qquad\qquad + a_{n,2n-1}\,x \end{cases}$$

sont des nombres rationnels, dont les dénominateurs ne contiennent que des facteurs premiers égaux à $n+1$ au plus.

Ce théorème est une conséquence immédiate de la propriété bien connue des coefficients numériques du polynome $\Phi_m(x)$, savoir que ces coefficients sont des nombres rationnels, dont les dénominateurs ne contiennent que des nombres premiers égaux à $m+1$ au plus,

Quant aux coefficients $a_{n,r}$ qui figurent au second membre de (9), introduisons, dans l'équation fonctionnelle (8), les expressions tirées de (9), puis cherchons, aux deux

membres de la formule ainsi obtenue, les coefficients de la puissance x^{2n-r-1}, il résulte la formule récursive générale

$$(10) \qquad \sum_{s=0}^{s=r} \binom{2n-r+s}{s+1} a_{n,\,r-s} = a_{n-1,\,r} + a_{n-1,\,r-1},$$

où il faut supposer $r \geq 1$, tandis que l'hypothèse $r = 0$ donnera

$$2n\,a_{n,\,0} = a_{n-1,\,0},$$

savoir

$$(11) \qquad a_{n,\,0} = \frac{1}{n!\,2^n}, \qquad n \geq 0.$$

On voit que la formule récursive (10), tirée de la définition (8), est très compliquée pour le calcul successif des coefficients $a_{n,\,r}$. Quant à la définition (5), elle donnera immédiatement

$$(12) \qquad a_{2n+1,\,4n+1} = 0;$$

parce que tous les termes qui figurent sous le signe sommatoire sont divisibles par x^2. De plus, nous aurons

$$\Phi'_{2n+1}(0) + 2n\,\Psi'_{4n}(0) = 0,$$

ce qui donnera, en vertu de (3),

$$(12\text{ bis}) \qquad a_{2n,\,4n-1} = \frac{(-1)^n B_n}{2n}.$$

Déterminons encore le coefficient $a_{n,\,2n-2}$. A cet effet, posons, dans (5), $p = 2n+1$, la puissance x^2 provient seulement des quatre termes

$$\Phi_{2n+2}(x), \quad -\Psi_2(x)\,\Phi_{2n+1}(x), \quad \Phi_1(x)\,\Psi_{4n}(x),$$
$$-(2n+1)\,\Psi_{4n+2}(x),$$

tandis que tous les autres termes sont divisibles par x^3, ce qui donnera, en vertu de (12 bis),

$$(13) \qquad a_{2n+1,\,4n} = \frac{(-1)^{n-1}(2n-1)\,B_n}{4n}.$$

Soit ensuite, dans (5), $p = 2n$, la puissance x^2 provient seulement des termes

$$2n\Psi_{4n}(x),\quad \Phi_{2n-s+1}(x)\Psi_{4s}(x),\qquad 1 \leqq s \leqq n-1,$$

tandis que tous les autres termes sont divisibles par x^3, de sorte que nous aurons, en vertu de (12 bis),

$$(13\text{ bis})\qquad a_{2n,\,4n-2} = \frac{(-1)^n}{2n}\cdot\sum_{s=1}^{s=n-1}\frac{B_s B_{n-s}}{2s},\qquad n \geqq 2.$$

Remarquons que les polynomes $\Phi_m(x)$ satisfont à l'équation fonctionnelle

$$(-1)^m\Phi_m(-x-1) = \Phi_m(x),$$

il résulte, en vertu de (8) et (12), le théorème:

IV. Soit n un positif entier, les polynomes $\Psi_{4n}(x)$ sont divisibles par $x(x+1)$, tandis que les polynomes $\Psi_{4n+2}(x)$ sont divisibles par $x^2(x+1)^2$.

De plus, introduisons, dans (5), $-x-1$ au lieu de x, il résulte

$$p\Psi_{2p}(-x-1) = \sum_{s=0}^{s=p-1}\Phi_{p-s-1}(x)\Psi_{2s}(-x-1),$$

d'où, en posant x égal au positif entier n,

$$(14)\qquad p\mathfrak{C}_{n+1}^{p} = \sum_{r=0}^{r=p-1}\mathfrak{C}_{n+1}^{r}S_{p-r}(n),$$

formule qui est analogue à celle de Newton concernant les nombres de Stirling de première espèce.

Soit maintenant p un nombre premier impair, on sait que les sommes de puissances $S_n(p)$ satisfont aux deux congruences

$$S_n(p) \equiv 0 \pmod{p},\qquad 1 \leqq n \leqq p-2$$

$$S_{2n+1}(p) \equiv 0 \pmod{p^2},\quad 1 \leqq n < \frac{p-3}{2},\ p \geqq 5,$$

ce qui donnera, en vertu de (2) et (14), le théorème:

V. Soit p un nombre premier impair, les nombres de Stirling de première et de seconde espèce satisfont aux congruences

$$(15) \qquad C_{p+1}^{n} \equiv \mathfrak{C}_{p+1}^{n} \equiv 0 \pmod{p},\ 1 \leq n \leq p-2,$$

$$(15\text{ bis}) \quad C_{p+1}^{2n+1} \equiv \mathfrak{C}_{p+1}^{2n+1} \equiv 0 \pmod{p^2},\ 1 \leq n \leq \frac{p-3}{2},\ p \geq 5.$$

La première des congruences (15) n'est autre chose que le lemme célèbre de LAGRANGE[1], appliqué dans sa démonstration, la première connue, du théorème de WILSON. Quant à la première des congruences (15 bis), WOLSTENHOLME[2] a connu le cas particulier qui correspond à $2n+1 = p-3$, conséquence immédiate des congruences de LAGRANGE du reste[3]. Ignorant qui est l'inventeur de la congruence générale en question, je me permets de remarquer que j'ai trouvé cette congruence, il y a trente ans à peu près.[4]

Remarquons ensuite que les formules (2) et (14) donnent les congruences modulo p:

$$\frac{2n}{p} C_{p+1}^{2n} \equiv \frac{2n}{p} \mathfrak{C}_{p+1}^{2n} \equiv -\frac{1}{p} S_{2n}(p)$$

$$\frac{2n+1}{p^2} C_{p+1}^{2n+1} \equiv \frac{1}{p^2} S_{2n+1}(p) - \frac{1}{2p} S_{2n}(p) + \frac{1}{2p} C_{p+1}^{2n},$$

$$\frac{2n+1}{p^2} \mathfrak{C}_{p+1}^{2n+1} \equiv \frac{1}{p^2} S_{2n+1}(p) + \frac{1}{2p} S_{2n}(p) + \frac{1}{2p} \mathfrak{C}_{p+1}^{2n},$$

il résulte, en vertu des résultats bien connus,

[1] Nouveaux Mémoires de l'Académie de Berlin, t. 2 (1771), p. 125—137; 1773.

[2] Quarterly Journal of Mathematics, t. 5, p, 35—39; 1862.

[3] Voir mon Traité élémentaire des Nombres de Bernoulli, article LXXXII.

[4] Nyt Tidsskrift for Mathematik, t. 4 B, p. 1—10; 1893.

$$\frac{1}{p} S_{2n}(p) \equiv (-1)^{n-1} B_n \pmod{p}$$

$$\frac{1}{p^2} S_{2n+1}(p) \equiv (-1)^{n-1}\left(n+\frac{1}{2}\right) B_n \pmod{p},$$

cet autre théorème, supplémentaire au précédent:

VI. Soit p un nombre premier, nous aurons les congruences modulo p:

$$\frac{1}{p} C_{p+1}^{2n} \equiv -\frac{1}{p} \mathfrak{C}_{p+1}^{2n} \equiv \frac{(-1)^n B_n}{2n}, \tag{16}$$

$$\begin{cases} \frac{1}{p^2} C_{p+1}^{2n+1} \equiv \frac{(-1)^{n-1}(2n-1) B_n}{4n}, \\ \frac{1}{p^2} \mathfrak{C}_{p+1}^{2n+1} \equiv \frac{(-1)^{n-1}(2n+1) B_n}{4n}, \end{cases} \tag{16 bis}$$

où il faut supposer

$$1 \leq n \leq \frac{p-3}{2}.$$

Ces résultats concernant les nombres C_{p+1}^m sont dus à M. GLAISHER.[1] Les deux congruences évidentes

$$C_{p+1}^p \equiv \mathfrak{C}_{p+1}^p \equiv 0 \pmod{p}$$

nous conduisent à l'étude des deux nombres C_p^{p-1} et $\mathfrak{C}_p^{p-1}$; remarquons que les nombres C_p^n, $\mathfrak{C}_p^n$ et $S_n(p-1)$ satisfont aux mêmes congruences que les C_{p+1}^n, $\mathfrak{C}_{p+1}^n$ et $S_n(p)$, les formules analogues à (2) et (14) donnent immédiatement

$$-(p-1) C_p^{p-1} \quad S_{p-1}(p-1) \quad -(p-1) \mathfrak{C}_p^{p-1} \pmod{p^2},$$

de sorte que la congruence évidente

$$S_{p-1}(p-1) \quad p-1 \equiv -1 \pmod{p}$$

donnera les deux autres

$$C_p^{p-1} \equiv -1 \pmod{p}, \quad \mathfrak{C}_p^{p-1} \equiv 1 \pmod{p}, \tag{17}$$

dont la première n'est autre chose que le théorème de WILSON.

[1] Quarterly Journal of Mathematics, t. 31, p. 321—353; 1900.

Posons ensuite

$$(18) \qquad C_p^{p-1} = -1 + p\,W_p, \qquad \mathfrak{S}_p^{p-1} = 1 + p\,V_p,$$

les W_p et les V_p sont des nombres entiers, et la congruence

$$S_p(p-1) \equiv (-1)^{n-1} B_n p \pmod{p^3}, \quad p = 2n,$$

donnera après un simple calcul

$$(19) \quad W_p \equiv V_p \equiv (-1)^{n-1} B_n \equiv \frac{1}{p} \equiv -1 \pmod{p}.$$

La congruence concernant le nombre W_p, le quotient de Wilson, est également due à M. Glaisher[1] et retrouvée par M. Lerch.[2]

Dans mes recherches sur les résidus quadratiques[3] j'ai donné des analogies aux congruences que nous venons d'étudier ici.

V.— Les polynomes de Stirling.

Le calcul direct des coefficients du polynome $\Psi_{2n}(x)$ étant très compliqué, nous avons à étudier, d'un autre point de vue, les polynomes en question.

A cet effet, revenons à l'équation fonctionnelle (8) de l'article précédent, savoir

$$(1) \quad \Psi_{2n}(x+1) = \Psi_{2n}(x) + (x+1)\,\Psi_{2n-2}(x), \qquad n \geq 1,$$

suppléée par les conditions ultérieures

$$(1\text{ bis}) \qquad \Psi_0(x) = 1, \quad \Psi_{2n}(0) = 0;$$

je dis que nous aurons le théorème général:

I. Soit $n \geq 1$, le polynome $\Psi_{2n}(x)$ est toujours divisible par la factorielle

$$(x+1)\,x\,(x-1)\,(x-2)\ldots(x-n+1),$$

[1] Quarterly Journal of Mathematics, t. 31, p. 321—353; 1900.

[2] Mathematische Annalen, t. 60, p. 488; 1905.

[3] Annales de l'École Normale, 3e série, t. 31, p. 198—200; 1914.

de sorte que nous aurons

$$(2)\quad \Psi_{2n}(x) = (x+1)\,x\,(x-1)\ldots(x-n+1)\,\psi_{n-1}(x),$$

où $\psi_{n-1}(x)$ est un polynome entier du degré $n-1$.

En effet, supposons $n \geqq 2$, la formule (1) donnera

$$\Psi_{2n}(1) = \Psi_{2n}(0) + \Psi_{2n-2}(0) = 0;$$

c'est-à-dire que $\Psi_{2n}(x)$ est, pour $n \geqq 2$, divisible par $x-1$; supposons ensuite $n > 3$, puis introduisons, dans (1), $x = 1$, nous aurons $\Psi_{2n}(2) = 0$, de sorte que $\Psi_{2n}(x)$ est, pour $n \geqq 3$, divisible par $x-2$. Continuons de cette manière, la conclusion de n à $n+1$ est évidente.

Dans ce qui suit nous désignons comme polynomes de Stirling la suite infinie de polynomes

$$\psi_0(x),\ \psi_1(x),\ \psi_2(x),\ \ldots,\ \psi_n(x),\ \ldots.$$

définis par les formules (2).

Soit particulièrement, dans (2), $n = 1$, on trouvera

$$(3)\qquad \psi_0(x) = \frac{1}{2}.$$

Introduisons ensuite, dans (1), les expressions tirées de (2), puis posons $n+1$ au lieu de n, il résulte les équations fonctionnelles

$$(4)\quad (x+2)\psi_n(x+1) = (x-n)\psi_n(x) + (x+1)\psi_{n-1}(x),$$

et il est très facile de démontrer le théorème fondamental:

II. Les équations fonctionnelles (4), suppléées par la valeur initiale $\psi_0(x) = \frac{1}{2}$, déterminent parfaitement la suite des polynomes de Stirling.

En effet, posons

$$(5)\quad \psi_n(x) = a_{n,0}x^n + a_{n,1}x^{n-1} + \ldots + a_{n,r}x^{n-r} + \ldots + a_{n,n},$$

l'équation fonctionnelle (4) donnera immédiatement

$$(6)\qquad (2n+2)\,a_{n,0} = a_{n-1,0},$$

et, pourvu que $p \geq 1$,

$$(6\text{ bis})\quad \left\{ \begin{array}{c} (2n-p+2)\,\alpha_{n,p} = \\ = \alpha_{n-1,p} + \alpha_{n-1,p-1} - \sum_{s=0}^{s=p-1} \left| \begin{matrix} n-s \\ p-s+1 \end{matrix} \right| + 2 \left| \begin{matrix} n-s \\ p-s \end{matrix} \right| \alpha_{n,s} \end{array} \right.$$

formules qui nous permettent de déterminer successivement les coefficients $\alpha_{n,p}$, pourvu que les $\alpha_{n-1,p}$ soient connus. Soit $p = n$, il faut, dans (6 bis), supprimer le terme illusoire $\alpha_{n-1,n}$; de plus, le premier coefficient binomial qui figure sous le signe sommatoire s'évanouira.

Remarquons que le théorème IV de l'article précédent donnera immédiatement cet autre:

III. Soit $n \geq 1$, les polynomes de Stirling à l'indice pair, savoir les $\psi_{2n}(x)$, sont divisibles par $x(x+1)$.

On voit que la détermination des coefficients du polynome $\psi_n(x)$ est très compliquée, aussi compliquée que la détermination des coefficients des polynomes $\Psi_{2n}(x)$. De plus, supposons connue la suite des polynomes de Stirling, nous avons à déterminer, en vertu de la formule (2), les $\Psi_{2n}(x)$, ce qui exige le développement d'une factorielle du rang n, savoir le problème que nous nous sommes proposé de résoudre. Néanmoins, les polynomes de Stirling jouent un rôle essentiel dans la théorie de la fonction exponentielle et du logarithme, nous le verrons bientôt.

Quant à ces applications transcendantes des polynomes de Stirling, nous avons besoin de plusieurs formules numériques qui sont des conséquences immédiates des définitions.

En premier lieu, il résulte, en vertu de (2) et des formules (7) de l'article précédent,

$$(7) \qquad C_{n+1}^{p} = \frac{(n+1)!}{(n-p)!}\,\psi_{p-1}(n), \qquad n \geq p > 0$$

$$(7\text{ bis}) \qquad \mathfrak{S}_{n+1}^{p} = \frac{(-1)^{p-1}(n-p)!}{(n-1)!}\,\psi_{p-1}(-n-1), \qquad p > 0.$$

Quant aux coefficients $\alpha_{n,r}$, la formule (6) donnera immédiatement

$$(8) \qquad \alpha_{n,0} = \frac{1}{(n+1)!\,2^{n+1}},$$

formule qui est aussi une conséquence de l'expression de $\alpha_{n+1,0}$, indiquée dans la formule (11) de l'article précédent.

De plus, le théorème III donnera

$$(9) \qquad \psi_{2n}(0) = \alpha_{2n,2n} = 0,$$

tandis que les formules (12 bis) et (13) de l'article précédent donnent

$$(10) \qquad \psi_{2n-1}(0) = \alpha_{2n-1,2n-1} = \frac{(-1)^{n-1}B_n}{(2n)!}$$

$$(10\text{ bis}) \qquad \psi'_{2n}(0) = \alpha_{2n,2n-1} = \frac{(-1)^{n-1}(2n-1)B_n}{(2n)!\,4n}.$$

Dans l'article XI, nous avons à déterminer d'autres valeurs numériques qui se rattachent aux polynomes de STIRLING, ce qui nous conduira, dans l'article XIV, à une formule récursive très curieuse pour les nombres de BERNOULLI.

DEUXIÈME PARTIE

De la fonction exponentielle.

VI.—Nouvelles définitions des $\psi_n(x)$.

Les polynomes de STIRLING que nous venons d'introduire par une méthode élémentaire, savoir à l'aide des nombres de STIRLING, proviennent aussi comme coefficients de certaines séries de puissances élémentaires, propriétés des $\psi_n(x)$ qui sont essentielles dans nos recherches suivantes.

En premier lieu, je dis que nous aurons, quelle que soit le variable complexe x,

$$(1)\quad \left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-x-1} = 1+(x+1)\cdot\sum_{n=1}^{n=\infty}\psi_{n-1}(x)\,\alpha^n, \quad |\alpha|<2\pi.$$

Remarquons tout d'abord qu'il existe une série de puissances de la forme indiquée et ayant son rayon de convergence égal à 2π; de plus, les coefficients de cette série deviennent des polynomes entiers de x, d'un degré égal à l'indice. Supposons ensuite

$$|x| \leqq K,$$

où K désigne un nombre positif, aussi grand qu'on le veut, je dis qu'il existe un autre nombre positif k, tel que la série qui figure au second membre de (1) est uniformément convergente par rapport à x, pourvu que

$$|\alpha| \leqq k.$$

A cet effet, posons pour abréger

$$f(\alpha) = \frac{\alpha}{2!} + \frac{\alpha^2}{3!} + \frac{\alpha^3}{4!} + \dots, \tag{2}$$

nous aurons évidemment

$$\frac{e^\alpha - 1}{\alpha} = f(\alpha) + 1,$$

de sorte que la formule (1) donnera

$$\left(\frac{e^\alpha - 1}{\alpha}\right)^{-x-1} = 1 + (x+1) \cdot \sum_{n=1}^{n=\infty} (-1)^n \psi_{n-1}(x)\, \alpha^n, \tag{3}$$

de plus, nous aurons, en appliquant la formule binomiale

$$\left(\frac{e^\alpha - 1}{\alpha}\right)^{-x-1} = 1 + \sum_{n=1}^{n=\infty} (-1)^n \binom{x+n}{n} (f(\alpha))^n. \tag{4}$$

Posons ensuite, conformément à la définition (2),

$$(f(\alpha))^n = \alpha^n (q_{n,0} + q_{n,1}\alpha + \dots + q_{n,p}\alpha^p + \dots), \tag{5}$$

la série de puissances ainsi obtenue est toujours convergente, et les coefficients $q_{n,p}$ sont des nombres positifs et rationnels.

Soit maintenant K un nombre positif, aussi grand qu'on le veut, il existe un nombre positif k, tel que les deux séries

$$(1 - f(\alpha))^{-K-1} = 1 + (K+1) \cdot \sum_{n=1}^{n=\infty} G_{n-1}(K)\, \alpha^n$$

$$(1 - f(\alpha))^{-K-1} = 1 + \sum_{n=1}^{n=\infty} \binom{K+n}{n} (f(\alpha))^n$$

sont uniformément convergentes, pourvu que $|\alpha| \leq k$, de sorte que nous aurons, en vertu de (5), et en appliquant un théorème bien connu de WEIERSTRASS,

$$(K+1)\, G_n(K+1) = \sum_{s=0}^{s=n} \binom{K+n-s+1}{n-s+1} q_{n-s+1,\,s}.$$

Cela posé, soit $|x| \leq K$, nous aurons évidemment

$$\left|\binom{x+n}{n}\right| \leq \binom{K+n}{n}, \quad n > 0,$$

c'est-à-dire que la série qui figure au second membre de (4) est absolument convergente, pourvu que $|\alpha| < k$, de sorte que nous aurons, en vertu de (3),

$$(6) \quad (x+1)\,\psi_n(x) = \sum_{s=0}^{s=n} (-1)^s \binom{x+n-s+1}{n-s+1} q_{n-s+1,\,s}.$$

Remarquons maintenant que les coefficients $q_{2n,\,r}$ sont des nombres positifs, nous aurons par conséquent

$$\left|(x+1)\,\psi_n(x)\right| \leq (K+1)\,G_n(K);$$

c'est-à-dire que la série qui figure au second membre de (1) est absolument et uniformément convergente par rapport à x, pourvu que nous ayons à la fois

$$|x| \leq K, \quad |\alpha| \leq k.$$

Cela posé, multiplions par α^{-x-1} les deux membres de la formule (1), puis différentions par rapport à x, terme à terme, ce qui est permis, nous retrouvons, pour les coefficients $\psi_n(x)$, l'équation fonctionnelle (4) de l'article précédent; de plus, nous aurons évidemment

$$\psi_0(x) = \frac{1}{2}.$$

La seconde série de puissances en question, savoir

$$(7) \quad \left(\frac{\log(1-\alpha)}{-\alpha}\right)^x = 1 + x \cdot \sum_{n=0}^{n=\infty} \psi_n(x+n)\,\alpha^{n+1}, \quad |\alpha| < 1,$$

peut être démontrée par la même méthode que la précédente.

Différentions par rapport à x les deux membres de la formule (7), puis posons $x = 0$, il résulte

$$(8) \qquad \log\left(\frac{-\alpha}{\log(1-\alpha)}\right) = \sum_{n=0}^{n=\infty} \psi_n(n)\,\alpha^{n+1}, \qquad |\alpha| < 1;$$

il est très facile d'exprimer, à l'aide des nombres connus, le coefficient général de cette dernière série de puissances.[1]

Ubbo Meyer[2] a étudié, le premier que je sache, les deux séries de puissances qui figurent aux seconds membres des formules (1) et (7), et, sans introduire les fonctions $\psi(x)$, il a vu la dépendance entre les deux suites de coefficients. Plus tard, Cayley[3] a étudié la série de puissances (8) et celles obtenues de (7), en différentiant r fois par rapport à x, puis posant $x = 0$, cependant il n'a pas trouvé la loi générale pour la formation des coefficients des séries de puissances ainsi obtenues.

Les dérivées, regardées par Cayley, et plusieurs autres du même genre sont contenues, comme des cas particuliers, dans les deux suivantes

$$(9) \qquad n!\,(x+1)\,\psi_{n-1}(x) = D_\alpha^n\left[\left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-x-1}\right]_{\alpha=0}$$

$$(10) \qquad n!\,x\,\psi_{n-1}(x+n-1) = D_\alpha^n\left[\left(\frac{\log(1-\alpha)}{-\alpha}\right)^{x}\right]_{\alpha=0},$$

tirées directement des formules (1) et (7).

Posons, dans (9) et (10), $x = p$ respectivement $x = -p$, où p désigne un positif entier, il résulte, en vertu des formules (7) et (7 bis) de l'article IV,

$$(11) \qquad C_{p+1}^n = \binom{p}{n} D_\alpha^n\left[\left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-p-1}\right]_{\alpha=0}, \qquad p \geqq n,$$

$$(12) \qquad \mathfrak{S}_{p-n+1}^n = (-1)^n\binom{p-1}{n} D_\alpha^n\left[\left(\frac{\log(1-\alpha)}{-\alpha}\right)^{-p}\right]_{\alpha=0}, \; p \geqq n+1;$$

[1] Voir mon Handbuch der Theorie der Gammafunktion, p. 77; Leipsic 1906.

[2] Archiv de Grunert, t. 9, p. 101—112; 1847.

[3] Quarterly Journal, t. 3, p. 366—369; 1860.

la première de ces formules est due à SCHLÖMILCH.[1] Ces deux dérivées connues, il résulte, en vertu de (1) et (7), les séries de puissances

$$(13)\quad \frac{p!}{(1-e^{-x})^{p+1}} = \sum_{s=0}^{s=p}(-1)^s(p-s)!\,\frac{C^s_{p-1}}{x^{p-s+1}} + (p+1)!\sum_{s=0}^{s=\infty}\psi_{p+s}(p)\,x^s$$

$$(14)\quad \frac{p!}{\left(\log(1+x)\right)^{p+1}} = \sum_{s=0}^{s=p}(-1)^s(p-s)!\,\frac{\mathfrak{S}^s_{p-s+2}}{x^{p-s+1}} + (p-1)!\sum_{s=0}^{s=\infty}(-1)^{p+s}\psi_{p+s}(s-1)\,x^s$$

dont les rayons de convergence sont égaux à 2π respectivement à 1; la première partie de la série de puissances qui figure au second membre de (13) est également due à SCHLÖMILCH[2] et plus tard étudiée par SYLVESTER.[3]

Les deux formules (1) et (7) donnent de même

$$(15)\qquad \mathfrak{S}^n_{p+1} = (-1)^n\binom{p+n}{n}D^n_\alpha\left[\left(\frac{1-e^{-\alpha}}{\alpha}\right)^p\right]_{\alpha=0}$$

$$(16)\qquad C^n_{p+n} = \binom{p+n}{n}D^n_\alpha\left[\left(\frac{\log(1-\alpha)}{-\alpha}\right)^p\right]_{\alpha=0},$$

où p désigne un positif entier, de sorte que nous aurons les deux séries de puissances très connues

$$(17)\qquad \frac{\left(\log(1+x)\right)^p}{p!} = \sum_{s=0}^{s=\infty}(-1)^s\,\frac{C^s_{p+s}\,x^{p+s}}{(p+s)!}$$

$$(18)\qquad \frac{(e^x-1)^p}{p!} = \sum_{s=0}^{s=\infty}\frac{\mathfrak{S}^s_{p+1}\,x^{p+s}}{(p+s)!}.$$

[1] Archiv de Grunert, t. 9, p. 333; 1847.

[2] Journal de Crelle, t. 44, p. 350; 1852.

[3] Voir le Jahrbuch über die Fortschritte der Mathematik, t. 15, p. 195; 1883.

dont la première a le rayon de convergence 1, tandis que la seconde est toujours convergente.

Pour mettre en pleine lumière le rôle fondamental que jouent les nombres de STIRLING, pour la fonction exponentielle et pour le logarithme, nous citons les deux formules différentielles très connues

$$(19) \qquad D_x^n f(\log x) = \frac{1}{x^n} \cdot \sum_{s=0}^{s=n-1} (-1)^s C_n^s f^{(n-s)}(t)_{t=\log x}$$

$$(20) \qquad D_x^n f(e^x) = \sum_{s=0}^{s=n-1} e^{(n-s)x} \mathfrak{S}_{n-s+1}^s f^{(n-s)}(t)_{t=e^x}.$$

Nous avons encore, dans cet article, à développer une suite des formules qui sont une conséquence directe de la série de puissance (1) et qui jouent un rôle fondamental dans nos recherches suivantes.

En premier lieu, prenons pour point de départ l'identité évidente

$$e^{\alpha(x+y+1)} \left(\frac{e^\alpha - 1}{\alpha}\right)^{-x-1} = e^{\alpha y} \left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-x-1},$$

puis cherchons le coefficient de la puissance α^{n+1}, il résulte l'identité algébrique

$$(21) \quad \left\{ \begin{aligned} &\frac{(x+y+1)^{n+1}}{(n+1)!} - (x+1) \cdot \sum_{s=0}^{s=n} \frac{(-1)^s (x+y+1)^{n-s}}{(n-s)!} \psi_s(x) = \\ &\qquad = \frac{y^{n+1}}{(n+1)!} + (x+1) \cdot \sum_{s=0}^{s=n} \frac{y^{n-s} \psi_s(x)}{(n-s)!}, \end{aligned} \right.$$

où x et y sont des variables complexes quelconques.

En second lieu, l'identité

$$\left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-x-1} \left(\frac{1-e^{\alpha}}{\alpha}\right)^{y+1} = \left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-x-y-2}$$

donnera de même

$$(22)\quad \begin{cases} (x+y+2)\,\psi_n(x+y+1) = (x-1)\,\psi_n(x) - (y+1)\,\psi_n(y) + \\ \qquad + (x+1)(y+1)\cdot \sum_{s=0}^{s=n-1} \psi_s(x)\,\psi_{n-s-1}(y). \end{cases}$$

On voit du reste que cette dernière identité n'est pas caractéristique pour les polynomes de Stirling, parce qu'elle est valable pour les polynomes $\varphi_n(x)$ définis par le développement

$$\big(1-\alpha f(\alpha)\big)^{-x-1} = 1+(x+1)\cdot \sum_{n=0}^{n=\infty} \varphi_{n-1}(x)\,\alpha^n,$$

où $f(\alpha)$ est une fonction holomorphe aux environs du point $\alpha = 0$.

En dernier lieu, nous avons à appliquer les polynomes de Bernoulli définis par les deux équations fonctionnelles

$$B_n'(x) = B_{n-1}(x), \qquad B_n(x) - B_n(x-1) = \frac{x^{n-1}}{(n-1)!}, \quad n \geq 1,$$

et intimement liés aux polynomes $\Phi_n(x)$, appliqués dans l'article IV, parce que nous aurons

$$\Phi_n(x) = (n-1)!\,\big(B_n(x) - B_n(0)\big).$$

A cet effet, prenons pour point de départ la série de puissances

$$\frac{e^{\alpha x}\,\alpha}{1-e^{-\alpha}} = \sum_{n=0}^{n=\infty} B_n(x)\,\alpha^n, \qquad |\alpha| < 2\pi,$$

qui conduira immédiatement aux deux équations fonctionnelles susdites, puis appliquons l'identité

$$\frac{e^{\alpha x}\,\alpha}{1-e^{-\alpha}} \left(\frac{e^{\alpha}-1}{\alpha}\right)^{-x-y-1} = \left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-x-1} \left(\frac{e^{\alpha}-1}{\alpha}\right)^{-y-1},$$

nous aurons, en cherchant le coefficient de la puissance x^{n+1},

$$(23)\quad \begin{cases} B_{n+1}(x) - (x-y+1)\cdot \sum_{s=0}^{s=n} (-1)^s \psi_s(x+y)\, B_{n-s}(x) = \\ \quad = (x+1)\,\psi_n(x) - (-1)^n (y+1)\,\psi_n(y) - \\ \quad\quad - (x+1)(y+1)\cdot \sum_{s=0}^{s=n-1} (-1)^s \psi_s(y)\,\psi_{n-s-1}(x). \end{cases}$$

Quant aux polynomes de BERNOULLI, nous avons encore à appliquer cette autre identité évidente

$$\frac{e^{-\frac{(x+1)\alpha}{2}}\alpha}{1-e^{-\alpha}}\left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-x} = \left(\frac{e^{\frac{\alpha}{2}}-e^{-\frac{\alpha}{2}}}{\alpha}\right)^{x-1};$$

remarquons ensuite que la fonction qui figure au second membre est une fonction paire de α, nous aurons

$$B_{2n+1}\left(-\frac{x+1}{2}\right) + x\cdot \sum_{s=0}^{s=2n} \psi_s(x-1)\, B_{2n-s}\left(-\frac{x+1}{2}\right) = 0,$$

d'où, en posant $x+1$ au lieu de x,

$$(24)\quad B_{2n+1}\left(\frac{x}{2}\right) = (x+1)\cdot \sum_{s=0}^{s=2n} (-1)^s \psi_s(x)\, B_{2n-s}\left(\frac{x}{2}\right),$$

car nous aurons, quel que soit l'indice n,

$$B_n(-x-1) = (-1)^n B_n(x).$$

VII.—Sur une série d'Euler.

Pour donner des éclaircissements sur une série de puissances, regardée par EULER [1], SCHERCK [2] et WORPITZKY [3], nous avons à étudier la fonction

[1] Institutiones calculi differentialis, p. 485—486, Pétrograde 1755.
[2] Journal de Crelle, t. 4, p. 302; 1829.
[3] Journal de Crelle, t. 94, p. 225; 1883.

(1) $$y = \frac{1}{b + a e^x},$$

où a et b sont des constantes quelconques. Un calcul direct donnera

$$\begin{aligned} ab\,y^2 &= y' + ay \\ 2(ab)^2 y^3 &= y'' + 3a\,y' + 2a^2 y \\ 6(ab)^3 y^4 &= y''' + 6a\,y'' + 11a^2 y' + 6a^3 y, \end{aligned}$$

ce qui nous conduira à poser généralement

(2) $$n!\,(ab)^n y^{n+1} = \sum_{s=0}^{s=n} C_{n+1}^{s}\, a^s\, y^{(n-s)}.$$

Quant à la conclusion de n à $n+1$, supposons vraie la formule (2), puis différentions par rapport à x, il résulte

$$(n+1)!\,(ab)^n\, y^n (ab\,y^2 - y) = \sum_{s=0}^{s=n} C_{n+1}^{s}\, a^s\, y^{(n-s+1)},$$

et la formule récursive

$$C_{n+2}^{s} = C_{n+1}^{s} + (n+1)\, C_{n+1}^{s-1}$$

conduira immédiatement au but.

Inversement, nous aurons

$$\begin{aligned} y' &= ab\,y^2 - a\,y \\ y'' &= 2(ab)^2 y^3 - 3a(ab)\,y^2 + a^2 y \\ y''' &= 6(ab)^3 y^4 - 12a\,(ab)^2 y^3 + 7a^2 (ab)\,y^2 - a^3 y, \end{aligned}$$

ce qui conduira à la formule générale

(3) $$y^{(n)} = \sum_{s=0}^{s=n} (-1)^s\, (n-s)!\, \mathfrak{S}_{n\ s+2}^{s}\, a^s\, (ab)^{n-s}\, y^{n-s-1}.$$

En effet, supposons vraie la formule (3), puis différentions par rapport à x, il résulte

$$y^{(n+1)} = \sum_{s=0}^{s=n} (-1)^s (n-s+1)!\, \mathfrak{S}_{n\ s+2}^{s}\, a^s (ab)^{n-s}\, y^{n-s}\, (ab\,y^2 - ay);$$

ordonnons ensuite, d'après des puissances descendantes de y, le second membre de cette dernière formule, la relation

$$\mathfrak{S}^{s}_{n-s+3} = \mathfrak{S}^{s}_{n-s+2} + (n-s+2)\,\mathfrak{S}^{s-1}_{n-s+3}$$

conduira immédiatement au but.

Étudions plus amplement le cas particulier $a = 1$, $b = -\alpha$, où α est un nombre complexe différent de l'unité, mais arbitraire du reste, il existe un développement de la forme

$$\frac{1}{\alpha - e^{x}} = \sum_{n=0}^{n=\infty} \frac{\varphi_n(\alpha)\,x^n}{n!\,(\alpha-1)^{n-1}}, \qquad \alpha \neq 1, \tag{4}$$

et nous aurons, en vertu de (3),

$$\varphi_n(\alpha) = \sum_{s=0}^{s=n} (-1)^s (n-s)!\,\mathfrak{S}^{s}_{n-s+2}\,\alpha^{n-s}(\alpha-1)^s, \tag{5}$$

tandis que la formule (2) donnera

$$n!\,\alpha^n = \sum_{s=0}^{s=n} C^{s}_{n+1}(\alpha-1)^s\,\varphi_{n-s}(\alpha). \tag{6}$$

Quant au polynome $\varphi_n(\alpha)$, défini par la formule (5), posons

$$\varphi_n(\alpha) = c_0 + c_1\alpha + c_2\alpha^2 + \ldots + c_r\alpha^r + \ldots,$$

nous aurons généralement, en vertu de (5),

$$c_r = \sum_{s=0}^{s=r} (-1)^s (r-s)! \binom{n-r+s}{s} \mathfrak{S}^{n-r+s}_{r-s+2}; \tag{7}$$

ordonnons ensuite, d'après les puissances $(r-s+1)^{n+1}$, l'expression indiquée pour c_r, le coefficient de cette puissance se présente sous la forme

$$(-1)^s \cdot \sum_{\mu=0}^{\mu=s} \frac{(r-\mu)!}{(r-\mu+1)!} \binom{n-r+\mu}{\mu} \binom{r-\mu+1}{s-\mu},$$

ou, ce qui est la même chose,

$$\frac{(-1)^s}{r-s+1}\sum_{\mu=0}^{\mu=s}\binom{n-r+\mu}{\mu}\binom{r-\mu}{s-\mu}=\frac{(-1)^s}{r-s+1}\binom{n+1}{s},$$

ce qui donnera finalement

$$c_r=\sum_{s=0}^{s=r}(-1)^s\binom{n+1}{s}(r-s+1)^n, \tag{8}$$

de sorte que nous aurons

$$c_n=\sum_{s=0}^{s=n}(-1)^s(n-s)!\,\mathfrak{S}_{n-s+2}^{s}=0. \tag{9}$$

Cela posé, il est évident que $\varphi_n(\alpha)$ est un polynome du degré $n-1$ par rapport à α, savoir

$$\varphi_n(\alpha)=c_0+c_1\alpha+c_2\alpha^2+\ldots+c_{n-1}\alpha^{n-1}; \tag{10}$$

on voit que l'expression générale des coefficients c_r n'est pas formellement la même que celle indiquée par Euler.[1]

Dans l'article IX nous avons à étudier, par une autre méthode, les polynomes $\varphi_n(\alpha)$.

VIII.—Deux séries de Lagrange.

Pour mettre en pleine lumière le rôle fondamental que les nombres de Stirling jouent dans la théorie de la fonction exponentielle et du logarithme, nous avons tout d'abord à étudier la série de Lagrange

$$F(x)=\sum_{n=0}^{n=\infty}A_n(e^x-1)^n. \tag{1}$$

A cet effet, supposons que la série de puissances

$$f(x)=\Sigma A_n x^n, \tag{2}$$

dont les coefficients sont les mêmes que ceux de la série

[1] Institutiones calculi differentialis, p. 485—486; Pétrograde 1755.

de Lagrange en question, ait le rayon de convergence $r > 0$; il est évident que le domaine de convergence de la série (1) est la partie du plan des x déterminée par la condition

$$|e^x - 1| < r, \tag{3}$$

d'où, en posant $x = \alpha + i\beta$, où α et β sont des variables réelles.

$$e^{2\alpha} - 2e^{\alpha}\cos\beta + 1 - r^2 < 0. \tag{4}$$

Cela posé, il est évident que le domaine de convergence ainsi déterminé est composé par une infinité de domaines convergents. En effet, soit $K_0(r)$ la partie du plan des x, définie par l'inegalité (4), suppléée par la condition ultérieure

$$-\pi \leqq \beta \leqq +\pi, \tag{5}$$

il est évident que l'on obtiendra un autre des domaines en question $K_p(r)$ en posant, dans $K_0(r)$, $\beta + 2p\pi$ à la place de β, où p désigne un nombre entier quelconque.

Dans ce qui suit nous désignons par $k_0(r)$ la frontière du domaine $K_0(r)$, savoir la courbe définie par l'équation en coordonnées rectangulaires

$$e^{2\alpha} - 2e^{\alpha}\cos\beta + 1 - r^2 = 0, \tag{6}$$

suppléée par la condition ultérieure (5); c'est-à-dire que la frontière $k_p(r)$ du domaine $K_p(r)$ peut être déduite de $k_0(r)$ en y remplaçant β par $\beta + 2p\pi$. De plus, nous désignons par $L_p(r)$ l'ensemble du domaine $K_p(r)$ et de sa frontière $k_p(r)$.

Quant à la courbe $k_0(r)$, que nous avons à étudier plus amplement, il résulte, en vertu de (6),

$$2(e^{2\alpha} - e^{\alpha}\cos\beta)\frac{d\alpha}{d\beta} + 2e^{\alpha}\sin\beta = 0;$$

c'est-à-dire que les valeurs maximums et minimums de α correspondent à $\beta = 0$, $\beta = \pm\pi$ respectivement.

Cela posé, nous avons à étudier séparément les trois cas suivants:

1^0 $r < 1$. Dans ce cas la courbe $k_0(r)$ est fermée et ses points d'intersection avec l'axe réelle et l'axe imaginaire sont déterminés par les valeurs

$$\alpha = \log(1 \pm r), \quad \beta = \pm \arccos \frac{2 - r}{2},$$

de sorte que nous aurons constamment

$$\log(1 - r) \leq \alpha \leq \log(1 + r).$$

Dans ce cas la série (1) représente généralement des fonctions différentes dans les domaines différents $K_n(r)$, à moins que la fonction $F(x)$ qu'il s'agit de développer n'ait la période additive $2\pi i$, savoir

$$F(x + 2\pi i) = F(x). \tag{7}$$

2^0 $r = 1$. La courbe $k_0(r)$ est une branche ouverte ayant les asymptotes $\beta = \pm \frac{\pi}{2}$, et dont les points d'intersection avec les axes sont déterminés par les valeurs

$$\alpha = \log 2, \quad \beta = \pm \frac{\pi}{3}.$$

Du reste, la série (1) a, dans ce cas, les mêmes propriétés que dans le cas précédent.

3^0 $r > 1$. Le domaine de convergence de la série de Lagrange en question est formé par une seule partie infiniment grande du plan des x, de sorte que la somme $F(x)$ de la série (1) est, dans ce cas, une fonction périodique de x, satisfaisant à la condition (7).

On voit que les valeurs maximums et minimums de α, qui correspondent à $\beta = 2p\pi$, respectivement $\beta = (2p + 1)\pi$, deviennent

$$\alpha = \log(r+1), \quad \alpha = \log(r-1).$$

Soit maintenant

$$(8) \qquad f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots$$

une fonction analytique aux environs de l'Origine, il existe toujours un développement de la forme

$$(9) \qquad f(x) = \sum_{s=0}^{s=\infty} A_s (e^x - 1)^s,$$

convergent dans un domaine $K_0(r)$, et dont les coefficients A_n sont à déterminer, d'après la formule de Lagrange[1], faussement attribuée à Bürmann[2], par les expressions

$$(10) \qquad \begin{cases} A_0 = a_0 \\ A_n = \dfrac{1}{n!} D_x^{n-1} \left[\left(\dfrac{x}{e^x - 1} \right)^n f'(x) \right]_{x=0}, \end{cases}$$

d'où, en vertu de la formule de Leibniz,

$$A_n = \frac{1}{n!} \cdot \sum_{s=0}^{s=n-1} \binom{n-1}{s} (n-s)!\, a_{n-s} D_x^s \left(\frac{x}{e^x - 1} \right)_{x=0}^n,$$

ce qui donnera finalement, en vertu de la formule (11) de l'article VI,

$$(11) \qquad \begin{cases} A_0 = a_0 \\ A_n = \displaystyle\sum_{s=0}^{s=n-1} \frac{(-1)^s (n-s)!}{n!} C_n^s\, a_{n-s}. \end{cases}$$

Inversement, prenons pour point de départ la série de Lagrange, qui figure au second membre de (9), cette série est uniformément convergente dans le domaine $L_0(r-\delta)$,

[1] Mémoires de l'Académie de Berlin, année 1768, p. 275.

[2] C. Ramus: Algebra og Functionslære, p. 108—112; Copenhague 1840, Schlömilch: Compendium der höheren Analysis, t. II, p. 102, 3e édition· Brunswick 1879.

où δ est une quantité positive arbitrairement petite, ce qui donnera, en vertu de la formule (18) de l'article VI,

$$(12)\qquad \begin{cases} a_0 = A_0 \\ a_n = \sum_{s=0}^{s=n-1} \frac{(n-s)!}{n!} \mathfrak{S}^s_{n-s+1} A_{n-s}. \end{cases}$$

En second lieu, nous avons à regarder cette autre série de Lagrange

$$(13)\qquad F(x) = \sum_{n=0}^{n=\infty} A_n \big(\log(1+x)\big)^n,$$

dont le domaine de convergence est déterminé par la condition

$$(14)\qquad \big|\log(1+x)\big| < r,$$

où r désigne le rayon de convergence de la série de puissances (2); on voit que le domaine en question contient toujours le point $x = 0$.

Supposons, dans ce cas, la série de puissances (8) développée comme suit

$$(15)\qquad f(x) = \sum_{s=0}^{s=\infty} A_s \big(\log(1+x)\big)^s,$$

nous aurons, en vertu de la formule de Lagrange

$$A_n = \frac{1}{n!} D_x^{n-1} \left[\left(\frac{x}{\log(1+x)}\right)^n f'(x)\right]_{x=0},$$

d'où, en appliquant la formule de Leibniz, puis introduisant les dérivées indiquées dans la formule (12) de l'article VI,

$$(16)\qquad \begin{cases} A_0 = a_0 \\ A_n = \sum_{s=0}^{s=n-1} \frac{(n-s)!}{n!} \mathfrak{S}^s_{n-s+1} a_{n-s}, \end{cases}$$

tandis que les formules inverses deviennent, en vertu de la formule (17) de l'article VI,

$$(17)\qquad \begin{cases} a_0 = A_0 \\ a_n = \sum_{s=0}^{s=n-1} \dfrac{(-1)^s (n-s)!}{n!} C_n^s A_{n-s}. \end{cases}$$

IX.—Applications diverses.

Soit par exemple

$$(1)\quad f(x) = \frac{x}{e^x - 1} = 1 - \frac{x}{2} + \sum_{n=1}^{n=\infty} \frac{(-1)^{n-1} B_n x^{2n}}{(2n)!}, \quad |x| < 2\pi,$$

où les B_n sont les nombres de Bernoulli, l'identité évidente

$$f(x) = \frac{\log\left(1 + (e^x - 1)\right)}{e^x - 1}$$

donnera, en vertu de la série de Mercator,

$$(2)\qquad \frac{x}{e^x - 1} = \sum_{n=0}^{n=\infty} \frac{(-1)^n}{n+1} (e^x - 1)^n,$$

développement qui n'est applicable que dans le domaine $K_0(1)$, tandis que la série de Lagrange en question représente d'autres fonctions dans les autres domaines $K_p(1)$.

Dans ce cas les formules générales (11) de l'article précédent donnent, après une simple réduction, la formule

$$(3)\qquad \frac{(n-1)!\,(n-1)}{2n+2} = \sum_{s=1}^{<\frac{n}{2}} (-1)^{s-1} C_n^{n-2s} B_s,$$

due à Slömilch [1] et retrouvée par A. Radicke [2]; dans

[1] Archiv de Grunert, t. 9, p. 334; 1847.

[2] Die Recursionsformeln für die Bernoullischen Zahlen, p. 15; Halle a. S. 1880.

l'article XIV nous avons à déduire, par une autre méthode, une généralisation très étendue de la formule (3).

Quant aux formules générales (12) de l'article précédent, nous aurons ici une expression explicite des nombres de BERNOULLI, savoir

$$(4) \qquad (-1)^{n-1} B_n = \sum_{s=0}^{s=2n-1} \frac{(-1)^s (2n-s)!}{2n \cdot s \mid 1} \mathfrak{S}^s_{2n-s+1}$$

et la formule récursive

$$(5) \qquad 0 = \sum_{s=0}^{s=2n} \frac{(-1)^s (2n-s+1)!}{2n-s+2} \mathfrak{S}^s_{2n-s+2}.$$

Étudions, comme second exemple, la fonction

$$(6) \qquad f(x) = \frac{1}{\alpha - e^x} = \frac{1}{\alpha - 1 - (e^x - 1)}, \qquad \alpha \neq 1,$$

il résulte la série de LAGRANGE

$$(7) \qquad \frac{1}{\alpha - e^x} = \sum_{n=0}^{n=\infty} \frac{(e^x-1)^n}{(\alpha-1)^{n+1}},$$

applicable dans l'ensemble de tous les domaines $K_p(r)$, où nous avons posé pour abréger $r = |\alpha - 1|$, car la fonction $f(x)$, dont il s'agit ici, admet la période additive $2\pi i$.

Posons, conformément à la formule (4) de l'article VII,

$$(8) \qquad \frac{1}{\alpha - e^x} = \sum_{n=0}^{n=\infty} \frac{\varphi_n(\alpha)\, x^n}{n!\,(\alpha-1)^{n+1}},$$

il résulte par conséquent, en vertu des formules générales (12) de l'article précédent, pour $\varphi_n(\alpha)$ l'expression

$$(9) \qquad \varphi_n(\alpha) = \sum_{s=0}^{s=n-1} (n-s)!\, \mathfrak{S}^s_{n-s+1} (\alpha-1)^s,$$

formellement différente de celle indiquée par la formule (5) de l'article VII, tandis que la formule inverse deviendra

$$(10) \qquad n! = \sum_{s=0}^{s=n-1} (-1)^s C_n^s \varphi_{n-s}(\alpha) (\alpha - 1)^s,$$

formule qui est très curieuse, comparée à la formule (6) de l'article VII.

Revenons maintenant à la formule (9), puis posons, conformément à l'expression (10) de l'article VII,

$$(11) \qquad \varphi_n(\alpha) = c_0 + c_1 \alpha + c_2 \alpha^2 + \ldots + c_{n-1} \alpha^{n-1},$$

nous aurons immédiatement

$$c_r = \sum_{s=0}^{s=n-r-1} (-1)^s (n-r-s)! \, \mathfrak{S}_{n-r-s+1}^{r+s} \binom{r+s}{s},$$

et le coefficient de la puissance $(n-r-\mu)^n$ qui figure dans cette expression deviendra

$$(-1)^\mu \sum_{s=0}^{s=n} \binom{r+s}{s} \binom{n-r-s}{\mu-s} = (-1)^\mu \binom{n+1}{\mu},$$

de sorte que nous aurons finalement

$$(12) \qquad c_r = \sum_{\mu=0}^{\mu=n-r-1} (-1)^\mu \binom{n+1}{\mu} (n-r-\mu)^n,$$

ce qui donnera, en vertu de la formule (8) de l'article VII,

$$c_r = c_{n-r-1};$$

c'est-à-dire que nous avons démontré la proposition, due à Scherk[1]:

La fonction $\varphi_n(x)$ est un polynome réciproque du degré $n-1$, savoir

$$(13) \qquad x^{n-1} \varphi_n\left(\frac{1}{x}\right) = \varphi_n(x).$$

Euler[2] a calculé les huit premiers des polynomes $\varphi_n(x)$, savoir

[1] Journal de Crelle, t. 4, p. 302, 1829.

[2] Institutiones calculi differentialis, p. 485—486, Pétrograde 1755.

$$\varphi_0(x) = 1$$
$$\varphi_1(x) = 1$$
$$\varphi_2(x) = x + 1$$
$$\varphi_3(x) = x^2 + 4x + 1$$
$$\varphi_4(x) = x^3 + 11x^2 + 11x + 1$$
$$\varphi_5(x) = x^4 + 26x^3 + 66x^2 + 26x + 1$$
$$\varphi_6(x) = x^5 + 57x^4 + 302x^3 + 302x^2 + 57x + 1$$
$$\varphi_7(x) = x^6 + 120x^5 + 1191x^4 + 2416x^3 + 1191x^2 + 120x + 1.$$

Désignons, comme ordinairement, par T_n les coefficients des tangentes, l'hypothèse $\alpha = -1$ donnera, en vertu de (8),

$$\varphi_{2n-1}(-1) = (-1)^{n-1} T_n$$
$$\varphi_0(-1) = \frac{1}{2}, \quad \varphi_{2n}(-1) = 0,$$

de sorte que nous aurons, en vertu de (9),

$$(14) \quad (-1)^{n-1} T_n = \sum_{s=0}^{s=2n-2} (-1)^s (2n-s-1)!\, 2^s \mathfrak{S}^s_{2n-s}$$

$$(15) \quad 0 = \sum_{s=0}^{s=2n-1} (-1)^s (2n-s)!\, 2^s \mathfrak{S}^s_{2n-s+1},$$

tandis que la formule (10) donnera de même

$$(16) \quad n! = \sum_{s=0}^{\leq \frac{n-1}{2}} (-1)^s 2^{n-2s-1} T_{s+1} C_n^{n-2s-1}.$$

La formule (14) est due à Laplace.[1]

Nous avons encore à développer la transcendante entière

$$f(x) = e^{\alpha x},$$

où α est une constante quelconque; les formules générales (11) du paragraphe précédent donnent, dans ce cas,

$$A_n = \sum_{s=0}^{s=n-1} \frac{(-1)^s \alpha^{n-s}}{n!} C_n^s = \binom{\alpha}{n},$$

[1] Mémoires de l'Académie des Sciences pour l'année 1777, p. 109—110 (1780).

de sorte que nous aurons la série de LAGRANGE, valable dans le domaine $K_0(1)$,

$$e^{\alpha x} = \sum_{n=0}^{n=\infty} \binom{\alpha}{n} (e^x - 1)^n , \tag{17}$$

évidente du reste. Quant aux formules (12) de l'article précédent, nous aurons, en changeant le signe de α,

$$\alpha^n = \sum_{s=0}^{s=n-1} (-1)^s \mathfrak{S}^s_{n-s+1}\, \omega_{n-s}(\alpha) , \tag{18}$$

savoir la formule de STIRLING, mentionnée dans l'article III; on connaît plusieurs démonstrations modernes de la formule (18), par exemple celle de CAUCHY.[1]

Cherchons encore, dans la formule (17), le coefficient de la puissance α^n, il résulte le développement

$$\frac{x^x}{n!} = \sum_{s=0}^{s=\infty} \frac{(-1)^s C^s_{n+s}}{(n+s)!} (e^x - 1)^{n+s} , \tag{19}$$

savoir la formule (17) de l'article VI, ce dernier développement est aussi valable dans le domaine $K_0(1)$.

Les deux séries de LAGRANGE qui figurent aux seconds membres des formules (17) et (19) sont intéressantes, parce qu'elles donnent des développements d'une transcendante entière d'après les puissances d'une autre transcendante entière, et les séries en question représentent des fonctions différentes dans chacun de leurs domaines de convergence, propriété analogue à celle des séries de FOURIER.

X.— D'autres applications.

Nous avons encore à étudier deux autres applications des formules générales développées dans l'article VIII,

[1] Résumés analytiques, p. 34; Turin 1833.

applications qui mettent en pleine lumière la nature très singulière des polynomes de STIRLING.

Combinons tout d'abord les deux séries de puissances (1) et (7) de l'article VI, posons dans la première de ces séries

$$\alpha = -\log(1+t),$$

il résulte

$$\left(\frac{\log(1+t)}{t}\right)^{x-1} = 1+(x+1)\cdot\sum_{n=0}^{n=\infty}(-1)^{n+1}\psi_n(x)\,(\log(1-t))^{n+1},$$

tandis que la dernière des séries susdites donnera

$$\left(\frac{\log(1+t)}{t}\right)^{x+1} = 1+(x+1)\cdot\sum_{n=0}^{n=\infty}(-1)^{n+1}\psi_n(x-n-1)\,t^{n+1}.$$

Cela posé, nous aurons, en vertu des formules générales (17) et (18) de l'article VI,

$$(1)\qquad \psi_n(x+n+1) = \sum_{s=0}^{s=n}\frac{(n-s+1)!}{(n+1)!}\,C^s_{n+1}\,\psi_{n-s}(x),$$

$$(1\text{ bis})\quad \psi_n(x) = \sum_{s=0}^{s=n}\frac{(-1)^s(n-s+1)!}{(n+1)!}\,\mathfrak{S}^s_{n-s+2}\,\psi_{n-s}(x+n-s+1),$$

ou, ce qui est la même chose,

$$(2)\quad \psi_n(x+n+1) = \psi_n(x)+\sum_{s=1}^{s=n}(n-s+1)\,\psi_{s-1}(n)\,\psi_{n-s}(x),$$

$$(2\text{ bis})\quad \left\{\begin{array}{l}\psi_n(x) = \psi_n(x+n+1)-\\ -\displaystyle\sum_{s=1}^{s=n}(n-s+1)\,\psi_{s-1}(-n+s-2)\,\psi_{n-s}(x+n-s+1);\end{array}\right.$$

dans les deux dernières formules il faut supposer naturellement $n > 1$.

Étudions ensuite le développement

$$(3)\qquad (1-\alpha)^{-x} = \sum_{n=0}^{n=\infty}\frac{\varphi_n(x)}{n!}\,(e^{\alpha}-1)^n,$$

ou, ce qui est la même chose,

$$(3\text{ bis}) \qquad (1 - \log(1-\alpha))^{-x} = \sum_{n=0}^{n=\infty} \frac{\varphi_n(x)}{n!}\alpha^n;$$

c'est-à-dire que la série (3) est applicable dans le domaine $K_0(1)$.

Différentions par rapport à α les deux membres de la formule (3 bis), puis multiplions par $\alpha+1$, il résulte pour les $\varphi_n(x)$ l'équation fonctionnelle

$$(4) \qquad x\varphi_n(x+1) = \varphi_{n+1}(x) + n\varphi_n(x),$$

tandis que les formules générales (11) et (12) de l'articleVIII donnent

$$(5) \qquad \varphi_n(x) = \sum_{s=0}^{s=n-1} (-1)^s C_n^s \omega_{n-s}(x)$$

$$(5\text{ bis}) \qquad \omega_n(x) = \sum_{s=0}^{s=n-1} \mathfrak{S}_{n-s+1}^{s} \varphi_{n-s}(x);$$

c'est-à-dire que $\varphi_n(x)$ est un polynome du degré n par rapport à x, et divisible par x, pourvu que $n > 0$.

Posons

$$(6) \quad \varphi_n(x) = a_{n,0}x^n + a_{n,1}x^{n-1} + \ldots + a_{n,r}x^{n-r} + \ldots + a_{n,n-1}x,$$

il résulte, en vertu de (5),

$$(7) \qquad a_{n,r} = \sum_{s=0}^{s=r} (-1)^s C_n^s C_{n-s}^{r-s},$$

ce qui donnera

$$(7\text{ bis}) \qquad a_{n,0} = 1, \quad a_{n,1} = 0.$$

Supposons maintenant, dans (7), $r \geq 2$, puis introduisons au second membre les polynomes de STIRLING, nous aurons

$$(8) \quad \left\{ \begin{array}{l} a_{n,r} = n(n-1)\ldots(n-r)\Big(\big(1+(-1)^r\big)\psi_{r-1}(n-1) + \\ + \sum_{s=1}^{s=r-1} (-1)^s (n-s)\psi_{s-1}(n-1)\psi_{r-s-1}(n-s-1)\Big), \end{array} \right.$$

ce qui nous conduira à étudier la fonction

$$(9)\quad \begin{cases} g_r(x) = (1-(-1)^r)\,\psi_r(x) \\ \qquad\qquad -\sum_{s=0}^{s=r-1}(-1)^s(x-s)\,\psi_s(x)\,\psi_{r-s-1}(x-s-1), \\ g_0(x) = 0, \end{cases}$$

parce que nous aurons, pour $r \geq 1$,

$$(10)\qquad a_{n,r} = n(n-1)\dots(n-r)\,g_{r-1}(n-1).$$

Quant aux polynomes $g_r(x)$, remarquons que les formules (4) et (5 bis) donnent les formules récursives

$$(11)\quad a_{n+1,r} - a_{n,r} + (r-1)\,a_{n,r-1} = \sum_{s=0}^{s=r-2}\binom{n-s}{r-s}a_{n,s}, \quad r \geq 2$$

$$(11\text{ bis})\qquad C_n^r = \sum_{s=0}^{s=r} \mathfrak{C}_{n-s+1}^{s}\,a_{n-s,\,r-s},$$

de sorte que nous aurons, en introduisant les polynomes $\psi_m(x)$ et $g_m(x)$,

$$(12)\quad \begin{cases} (x+2)\,g_r(x+1) - (x-r)\,g_r(x) + r\,g_{r-1}(x) = \\ \qquad = \dfrac{1}{(r+1)!} + \sum_{s=0}^{s=r-2}\dfrac{(x-s)\,g_s(x)}{(r-s)!} \end{cases}$$

$$(12\text{ bis})\quad \begin{cases} \psi_r(x) = g_r(x) + \\ + \sum_{s=0}^{s=r-1}(-1)^s(x-s)\,g_{r-s-1}(x-s)\,\psi_s(-x+s-1) + \\ \qquad + (-1)^r\psi_r(-x+r-1); \end{cases}$$

dans ces deux formules il faut supposer $r \geq 2$ respectivement $r \geq 1$.

On trouve par exemple

$$g_0(x) = 0, \quad g_1(x) = \frac{1}{3!}, \quad g_2(x) = -\frac{1}{4!}, \quad g_3(x) = \frac{5x+4}{5!\,3},$$

ce qui donnera

$$a_{n,0} = 1, \quad a_{n,1} = 0, \quad a_{n,2} = \binom{n}{3}, \quad a_{n,3} = -\binom{n}{4},$$

$$a_{n,4} = \binom{n}{5}\frac{5n-1}{3},$$

de sorte que nous aurons

$$\begin{aligned}
\varphi_0(x) &= 1 \\
\varphi_1(x) &= x \\
\varphi_2(x) &= x^2 \\
\varphi_3(x) &= x^3+x \\
\varphi_4(x) &= x^4+4x^2-x \\
\varphi_5(x) &= x^5+10x^3-5x^2+8x \\
\varphi_6(x) &= x^6+20x^4-15x^3+58x^2-26x.
\end{aligned}$$

Soit, dans (5), $x=-1$, on aura généralement

$$\varphi_n(-1) = (-1)^n(n-1)!, \tag{13}$$

formule qui peut servir comme contrôle des calculs précédents.

Quant au polynome $g_n(x)$, on voit que son degré est toujours inférieur à son indice. Or, une étude plus profonde de cette question nous conduira beaucoup trop loin ici.

En dernier lieu, posons

$$\left(\frac{1+e^{\alpha}}{2}\right)^{-x} = \sum_{n=0}^{n=\infty} \frac{(-1)^n \varphi_n(x)\,\alpha^n}{n!\,2^n}, \qquad |\alpha| < \frac{\pi}{2}, \tag{14}$$

l'identité évidente

$$\left(\frac{1+e^{\alpha}}{2}\right)^{-x} = \left(1-\frac{1-e^{\alpha}}{2}\right)^{-x}$$

donnera la série de Lagrange

$$\left(\frac{1+e^{\alpha}}{2}\right)^{-x} = \sum_{n=0}^{n=\infty} \frac{(-1)^n}{2^n}\binom{x+n-1}{n}(e^{\alpha}-1)^n, \tag{14 bis}$$

valable dans l'ensemble des domaines $K_p(2)$, parce que la fonction, dont il s'agit ici, admet la période additive $2\pi i$.

Cela posé, il résulte, en vertu des formules générales (11) et (12) du paragraphe VIII,

$$\varphi_n(x) = \sum_{s=0}^{s=n-1} (-2)^s \, \mathfrak{S}^s_{n-s+1} \, \omega_{n-s}(x) \tag{15}$$

$$\omega_n(x) = \sum_{s=0}^{s=n-1} 2^s \, C^s_n \, \varphi_{n-s}(x), \tag{15 bis}$$

tandis que nous aurons l'équation fonctionnelle

$$\varphi_{n+1}(x) = 2x\varphi_n(x) - x\varphi_n(x+1), \tag{16}$$

obtenue immédiatement en différentiant par rapport à α la formule (14), puis appliquant l'identité évidente $e^\alpha = (e^\alpha + 1) - 1$.

Posons ensuite, conformément à (15),

$$\varphi_n(x) = a_{n,0}x^n + a_{n,1}x^{n-1} + \ldots + a_{n,p}x^{n-p} + \ldots + a_{n,n-1}x, \tag{17}$$

il résulte, en vertu de (16),

$$\left\{ \begin{array}{l} a_{n+1,0} = a_{n,0} = 1 \\ a_{n+1,r} = a_{n,r} - \displaystyle\sum_{s=0}^{s=r-1} \binom{n-r}{r-s} a_{n,s}. \end{array} \right. \tag{18}$$

tandis que les formules (15) donnent respectivement

$$a_{n,r} = \sum_{s=0}^{s=r} (-2)^s \, \mathfrak{S}^s_{n-s-1} \, C^{r-s}_{n-s} \tag{19}$$

$$C^r_n = \sum_{s=0}^{s=r} 2^s \, C^s_n \, a_{n-s,\,r-s}. \tag{19 bis}$$

On voit que les coefficients $a_{n,r}$, dont il s'agit ici, sont de la même forme que ceux de l'exemple précédent, de sorte qu'il existe une expression analogue à celle indiquée dans la formule (10), mais le polynome $g_r(x)$ correspondant est toujours du degré r par rapport à x. On trouve ici

$$a_{n,0} = 1, \quad a_{n,1} = -\binom{n}{2}, \quad a_{n,2} = \binom{n}{3}\frac{3n-9}{4},$$

et les premiers des polynomes $\varphi_n(x)$ deviennent

$$\begin{aligned}
\varphi_0(x) &= 1 \\
\varphi_1(x) &= x \\
\varphi_2(x) &= x^2 - x \\
\varphi_3(x) &= x^3 - 3x^2 \\
\varphi_4(x) &= x^4 - 6x^3 + 3x^2 + 2x \\
\varphi_5(x) &= x^5 - 10x^4 + 15x^3 + 10x^2.
\end{aligned}$$

Remarquons que l'équation fonctionnelle (16) donnera

$$\varphi'_{n+1}(0) = -\varphi_n(1), \quad \varphi_n(2) = 2\varphi_n(1) - \varphi_{n+1}(1),$$

il résulte, en vertu de la série de puissances

$$\frac{1}{1+e^{\alpha}} = \frac{1}{2} + \sum_{n=1}^{n=\infty} \frac{(-1)^n T_n \alpha^{2n-1}}{(2n-1)!\, 2^{2n}}, \quad |\alpha| < \pi,$$

où les T_n sont les coefficients des tangentes, les valeurs numériques

$$(20) \quad \begin{cases} \varphi'_{2n}(0) = (-1)^n T_n, & \varphi'_{2n+1}(0) = 0 \\ \varphi_{2n}(1) = 0, & \varphi_{2n+1}(1) = (-1)^n T_{n+1} \\ \varphi_{2n}(2) = (-1)^{n-1} T_{n+1}, & \varphi_{2n+1}(2) = (-1)^n 2T_{n+1}. \end{cases}$$

On voit que ces formules donnent, en vertu de (15), des expressions explicites des coefficients des tangentes, expressions que j'ai étudiées dans une autre occasion.[1]

Posons, dans (14), $x = -1$, il résulte cette autre valeur numérique

$$(21) \qquad \varphi_n(-1) = (-1)^n 2^{n-1}.$$

[1] Annali di matematica (3) t. 19, p. 197—200; 1912.

TROISIÈME PARTIE

Les polynomes de Stirling.

XI.—Évaluation des valeurs numériques.

Dans nos études suivantes sur les polynomes de STIRLING nous appliquons indistinctement la définition élémentaire, à l'aide des nombres de STIRLING, et la définition transcendante, à l'aide des deux séries de puissances (1) et (7) de l'article VI.

Les recherches plus profondes sur les polynomes $\psi_n(x)$ exigent des valeurs numériques qui sont des conséquences immédiates de la définition élémentaire.

En premier lieu, la formule (7) de l'article V donnera

$$(1) \qquad \psi_n(n+1) = \frac{1}{n+2}$$

$$(1\text{ bis}) \qquad \psi_n(n+2) = \frac{1}{n+3}\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n+2}\right),$$

tandis que la formule (7 bis) du même article nous conduira à ces deux résultats

$$(2) \qquad \psi_n(-2) = \frac{(-1)^n}{(n+2)!}$$

$$(2\text{ bis}) \qquad \psi_n(-3) = \frac{(-1)^n(2^{n+2}-1)}{(n+3)!}.$$

En second lieu, il est évident que l'équation fonctionnelle

$$(3) \quad (x+2)\,\psi_n(x+1) = (x-n)\,\psi_n(x) + (x+1)\,\psi_{n-1}(x)$$

nous permet de déterminer d'autres valeurs numériques.

En effet, prenons pour point de départ les expressions

$$(4) \qquad \psi_{2n-1}(0) = \frac{(-1)^{n-1}B_n}{(2n)!}, \quad \psi_{2n}(0) = 0, \quad n \geq 1,$$

indiquées dans les formules (9) et (10) de l'article V, nous aurons, en posant, dans (3), $x = -1$,

$$(5) \qquad \psi_{2n-1}(-1) = \frac{(-1)^{n} B_n}{(2n)!\,2n}, \quad \psi_{2n}(-1) = 0, \quad n \geq 1,$$

tandis que la valeur

$$(6) \qquad \psi'_{2n}(0) = \frac{(-1)^{n-1}(2n-1)B_n}{(2n)!\,4n}, \qquad n \geq 1,$$

indiquée dans la formule (10 bis) de l'article V, donnera, en vertu de la relation

$$\psi'_{2n}(0) = -(2n+1)\,\psi'_{2n}(-1) + \psi_{2n-1}(-1),$$

tirée directement de (3),

$$(6 \text{ bis}) \qquad \psi'_{2n}(-1) = \frac{(-1)^{n-1}B_n}{(2n)!\,4n}.$$

Soit ensuite, dans (3), $x = 0$, la formule (4) donnera

$$(7) \qquad \psi_{2n-1}(1) = \frac{(-1)^{n}(2n-1)B_n}{(2n)!\,2}, \qquad n \geq 2,$$

$$(7 \text{ bis}) \qquad \psi_{2n}(1) = \frac{(-1)^{n-1}B_n}{(2n)!\,2}, \qquad n \geq 1,$$

tandis que nous aurons

$$\psi_0(1) = \frac{1}{2}, \quad \psi_1(1) = \frac{5}{24}.$$

Posons, encore, dans (3), $x = 1$, nous aurons

$$3\psi_n(2) = -(n-1)\,\psi_n(1) + 2\,\psi_{n-1}(1),$$

d'où, en vertu des formules (7),

$$(8) \qquad \psi_{2n}(2) = \psi_{2n-1}(1) = \frac{(-1)^{n}(2n-1)B_n}{(2n)!\,2}$$

(8 bis) $$\psi_{2n\ 1}(2) = \frac{(-1)^{n-1}(2n-1)(2n-2)B_n}{2n!\,6} + \frac{(-1)^n B_{n-1}}{(2n-2)!\,3},$$

où il faut supposer $n \geq 2$. Quant aux valeurs exclues qui correspondent à $n = 0, 1, 2$, un calcul direct donnera

$$\psi_0(2) = \frac{1}{2}, \quad \psi_1(2) = \frac{1}{3}, \quad \psi_2(2) = \frac{1}{8}.$$

Étudions maintenant les coefficients $\alpha_{n,r}$ du polynome $\psi_n(x)$, savoir

(9) $$\psi_n(x) = \alpha_{n,0}x^n + \alpha_{n,1}x^{n-1} + \ldots + \alpha_{n,r}x^{n-r} + \ldots + \alpha_{n,n},$$

nous avons, déjà dans l'article V, indiqué les résultats

(10) $$\alpha_{n,0} = \frac{1}{(n+1)!\,2^{n+1}},$$

(11) $$\alpha_{2n-1,\,2n-1} = \frac{(-1)^{n-1}B_n}{(2n)!}, \qquad \alpha_{2n,\,2n} = 0, \qquad n \geq 1$$

(12) $$\alpha_{2n,\,2n-1} = \frac{(-1)^{n-1}(2n-1)B_n}{(2n)!\,4n}, \qquad n \geq 1.$$

Déterminons encore la valeur de

$$\alpha_{2n-1,\,2n-2} = \psi'_{2n-1}(0);$$

à cet effet, prenons pour point de départ la formule

$$\Psi_{4n}(x) = (x+1)\,x\,(x-1)\,(x-2)\ldots(x-2n+1)\,\psi_{2n-1}(x),$$

tirée directement de la définition de $\psi_{n\ 1}(x)$, savoir la formule (2) de l'article V. Appliquons ensuite le développement

$$(x+1)\,x\,(x-1)\ldots(x-2n+1) =$$
$$= -C_{2n}^{2n-1}x + (C_{2n}^{2n-2} - C_{2n}^{2n-1})x^2 + \ldots,$$

il résulte, en vertu de (9),

$$\Psi_{4n}(x) = -C_{2n}^{2n-1}\alpha_{2n-1,\,2n-1}x +$$
$$+ \left((C_{2n}^{2n-2} - C_{2n}^{2n-1})\,\alpha_{2n-1,\,2n-1} - C_{2n}^{2n-1}\alpha_{2n-1,\,2n-2}\right)x^2 + \ldots,$$

de sorte que les deux derniers coefficients de l'expression

$$\Psi_{4n}(x) = a_{2n,\,0}\,x^{4n} + a_{2n,\,1}\,x^{4n-1} + \ldots + a_{2n,\,4n-2}x^2 + a_{2n,\,4n-1}\,x$$

deviennent

$$a_{2n,\,4n-1} = (2n-1)!\,a_{2n-1,\,2n-1}$$

$$a_{2n,\,4n-2} = \left(C_{2n}^{2n-2} - (2n-1)!\right) a_{2n-1,\,2n-1} - (2n-1)!\,a_{2n-1,\,2n-2}.$$

Cela posé, nous aurons évidemment

$$2n\,a_{2n,\,4n-2} = \frac{(-1)^{n-1}C_{2n}^{2n-2}B_n}{(2n-1)!} + (-1)^n B_n - (2n)!\,a_{2n-1,\,2n-2},$$

d'où, en vertu de la formule (13 bis) de l'article IV,

$$(13)\quad (-1)^{n-1}(2n)!\,a_{2n-1,\,2n-2} = \frac{B_n\,C_{2n}^{2n-2}}{(2n-1)!} - B_n + \sum_{s=1}^{s=n-1}\frac{B_s\,B_{n-s}}{2s},$$

ou, ce qui est la même chose,

$$(13\text{ bis})\quad \left\{\begin{array}{c} (-1)^{n-1}(2n)!\,\psi'_{2n-1}(0) = \\ = \left(\dfrac{1}{2} + \dfrac{1}{3} + \dfrac{1}{4} + \ldots + \dfrac{1}{2n-1}\right)B_n + \displaystyle\sum_{s=1}^{s=n-1}\frac{B_s\,B_{n-s}}{2s}; \end{array}\right.$$

dans ces deux formules il faut supposer naturellement $n \geq 2$.

L'expression ainsi trouvée pour $a_{2n-1,\,2n-2}$ montre clairement que la nature des coefficients $a_{n,\,r}$ est très compliquée; dans l'article XIV nous avons à étudier, d'un autre point de vue, les coefficients $a_{2n-1,\,2n-2}$.

Quant aux coefficients $a_{n,\,r}$, remarquons encore que les expressions $\psi_n(1) \pm \psi_n(-1)$, formées en vertu des formules (5) et (7), donnent immédiatement

$$a_{2n,\,0} + a_{2n,\,2} + \ldots + a_{2n,\,2n} = \frac{(-1)^{n-1}B_n}{(2n)!\,4}$$

$$a_{2n,\,1} + a_{2n,\,3} + \ldots + a_{2n,\,2n-1} = \frac{(-1)^{n-1}B_n}{(2n!)\,4}$$

$$a_{2n-1,\,0} + a_{2n-1,\,2} + \ldots + a_{2n-1,\,2n-2} = \frac{(-1)^n(2n-1)(n-1)B_n}{(2n)!\,4n}$$

$$\alpha_{2n-1,1}+\alpha_{2n-1,3}+\ldots+\alpha_{2n-1,2n-1}=\frac{(-1)^n(2n-1)(n+1)B_n}{(2n)!\,4n}.$$

Il saute aux yeux que les valeurs numériques, que nous venons de démontrer, montrent clairement que les coefficients $\alpha_{n,r}$ sont intimement liés aux nombres de Bernoulli, mais qu'ils sont d'une nature beaucoup plus compliquée.

Revenons maintenant à l'équation fonctionnelle (3), puis posons $x+n$ au lieu de x, il résulte

$$(x+n+2)\psi_n(x+n+1)=x\psi_n(x+n)+(x+n+1)\psi_{n-1}(x+n),$$

formule récursive qui donnera immédiatement

$$(14)\quad (x+n+2)\,\psi_n(x+n+1)=1+x\cdot\sum_{s=0}^{s=n}\psi_s(x+s);$$

posons, dans cette dernière formule $x=0$, puis $x=1$, nous retrouvons les formules (2).

XII.— Applications des nombres de Stirling.

Le calcul direct des polynomes de Stirling étant très compliqué, nous avons à développer une suite de formules récursives qui nous permettent de calculer successivement les polynomes en question.

En premier lieu, remarquons que l'introduction des polynomes $\Psi_{2r}(x)$ nous permet de généraliser beaucoup les formules numériques (3) de l'article II. A cet effet, écrivons, en vertu des formules (7) de l'article IV, la première des formules en question sous la forme

$$\sum_{s=0}^{s=r}(-1)^s\,\Psi_{2r-2s}(n+p)\,\Psi_{2s}(-n-1)=$$
$$=\sum_{s=0}^{s=r}\binom{p-s}{r-s}(n-1)^{r-s}\,\Psi_{2s}(p-1),$$

je dis que nous aurons tout d'abord l'identité algébrique

$$(1)\quad \left\{ \begin{array}{l} \displaystyle\sum_{s=0}^{s=r} (-1)^s \Psi_{2r-2s}(y)\, \Psi_{2s}(-n-1) = \\ \displaystyle = \sum_{s=0}^{s=r} \binom{y-n-s}{r-s} (n+1)^{r-s}\, \Psi_{2s}(y-n-1), \end{array} \right.$$

où y désigne une variable complexe quelconque, tandis que n est un positif entier fixe, mais arbitraire du reste.

En effet, l'équation algébrique (1), du degré $2r$ par rapport à y au plus, admet comme racines toutes les valeurs de la forme $y = n+p$, où p est un nombre entier plus grand que r.

Cela posé, nous aurons l'identité algébrique beaucoup plus générale

$$(2)\quad \left\{ \begin{array}{l} \displaystyle\sum_{s=0}^{s=r} (-1)^s \Psi_{2s}(x)\, \Psi_{2r-2s}(y) = \\ \displaystyle = \sum_{s=0}^{s=r} (-1)^{r-s} \binom{x+y-s+1}{r-s} x^{r-s}\, \Psi_{2s}(x+y) \end{array} \right.$$

où x est une nouvelle variable complexe quelconque. En effet, soit y un nombre fixe, l'équation (2) admet comme racines $x = -n-1$, où n désigne un positif entier quelconque.

Soit maintenant, dans (2), $y = 0$, le premier membre se réduit à son dernier terme, de sorte que nous aurons la formule récursive

$$(3)\quad \left(1-(-1)^r\right) \Psi_{2r}(x) = \sum_{s=0}^{s=r-1} (-1)^s \binom{x-s+1}{r-s} x^{r-s}\, \Psi_{2s}(x),$$

d'où, en introduisant, en vertu de la définition (2) de l'article V, les polynomes de STIRLING, puis posant $r+1$ au lieu de r,

$$(4)\quad (1+(-1)^r)(x-r)\psi_r(x) = \frac{x^{r+1}}{(r-1)!} - \sum_{s=0}^{s=r-1}(-1)^s\frac{(x-s)x^{r-s}}{(r-s)!}\psi_s(x),$$

où il faut supposer naturellement $r \geq 1$.

Posons ensuite, dans (2), $y = x$, puis supposons impair le nombre r, le premier membre de la formule en question s'évanouira, et nous aurons, en remplaçant r par $2r+1$ et x par $\frac{x}{2}$,

$$(5)\quad \Psi_{4r+2}(x) = \sum_{s=0}^{s=2r}(-1)^s\binom{x-2r-s+1}{s+1}\left(\frac{x}{2}\right)^{s+1}\Psi_{4r-2s}(x),$$

ce qui donnera, pour les polynomes de Stirling,

$$(6)\quad \begin{cases} (x-2r)\psi_{2r}(x) = \\ = \dfrac{x^{2r+1}}{(2r+1)!\,2^{2r+1}} - \displaystyle\sum_{s=0}^{s=2r-1}\frac{(-1)^s(x-2r+s+1)}{(s+1)!}\left(\frac{x}{2}\right)^{s+1}\psi_{2r-s-1}(x). \end{cases}$$

Quant à la formule (3 bis) de l'article II, savoir

$$\sum_{s=0}^{s=r}(-1)^s\Psi_{2s}(n)\,\Psi_{2r-2s}(-n-p-1) =$$

$$= \sum_{s=0}^{s=r}\binom{p+r-1}{r-s}(n+1)^{r-s}\Psi_{2s}(-p),$$

nous posons

$$n = x, \quad -n-p-1 = y, \quad -p = x-y-1,$$

et une simple réduction donnera l'identité algébrique

$$(7)\quad \begin{cases} \displaystyle\sum_{s=0}^{s=r}(-1)^s\Psi_{2s}(x)\,\Psi_{2r-2s}(y) = \\ = \displaystyle\sum_{s=0}^{s=r}(-1)^{r-s}\binom{x+y-s+1}{r-s}(x+1)^{r-s}\Psi_{2s}(x+y+1), \end{cases}$$

où x et y sont des variables complexes quelconques.

Posons, en premier lieu, $y = -1$, il résulte

$$(8)\quad (1-(-1)^r)\,\Psi_{2r}(x) = \sum_{s=0}^{s=r-1}(-1)^s\binom{x-s}{r-s}(x+1)^{r-s}\,\Psi_{2s}(x),$$

d'où, en introduisant les polynomes de STIRLING, puis remplaçant r par $r+1$,

$$(9)\quad \left\{\begin{array}{l} (1+(-1)^r)\,\psi_r(x) = \\ \\ = \dfrac{(x+1)^r}{(r+1)!} - \displaystyle\sum_{s=0}^{s=r-1}\frac{(-1)^s(x+1)^{r-s}}{(r-s)!}\,\psi_s(x), \end{array}\right.$$

où il faut supposer naturellement $r \geq 1$.

Soit ensuite, dans (7), r un nombre impair, l'hypothèse $y = x$ donnera

$$0 = \sum_{s=0}^{s=r}(-1)^{r-s}\binom{2x-s+1}{r-s}(x+1)^{r-s}\,\Psi_{2s}(2x+1),$$

d'où, en introduisant $\frac{x}{2}-\frac{1}{2}$ au lieu de x, $2r+1$ au lieu de r,

$$(10)\quad \Psi_{4r-2}(x) = \sum_{s=0}^{s=2r}(-1)^s\binom{x-s}{2r-s+1}\left(\frac{x+1}{2}\right)^{2r-s+1}\Psi_{2s}(x),$$

de sorte que nous aurons pour les polynomes de STIRLING

$$(11)\quad \psi_{2r}(x) = \frac{(x+1)^{2r}}{(2r+1)!\,2^{2r-1}} - \sum_{s=0}^{s=2r-1}\frac{(-1)^s}{(2r-s)!}\left(\frac{x+1}{2}\right)^{2r-s}\psi_s(x).$$

Il est curieux, ce me semble, que les formules récursives, tirées de l'identité (7), soient des conséquences immédiates de la théorie des polynomes symétriques, nous le verrons dans l'article qui suit. De plus, posons, dans l'identité (21) de l'article VI, $y = 0$, nous retrouvons précisément la formule (9).

Quant aux quatre formules récursives, ainsi obtenues pour les polynomes de STIRLING, il est évident qu'elles nous permettent de calculer les $\psi_{2r}(x)$ à l'aide de $\psi_s(x)$.

où $s < 2r$, tandis que ces formules ne présentent aucun moyen pour le calcul des $\psi_{2r+1}(x)$ à l'aide des polynomes précédents. Pour obtenir des formules récursives de ce genre nous posons, dans (2), $x = -1$, puis introduisons $x+1$ au lieu de y, ce qui donnera

$$\Psi_{2r}(x+1) - \Psi_{2r}(x) = \sum_{s=0}^{s=r-1} \binom{x-s+1}{r-s} \Psi_{2s}(x).$$

Appliquons ensuite l'équation fonctionnelle (8) de l'article IV, puis remplaçons r par $r+1$, il résulte finalement

$$r\Psi_{2r}(x) = \sum_{s=0}^{s=r-1} \binom{x-s+1}{r-s+1} \Psi_{2s}(x), \tag{12}$$

ce qui donnera pour les polynomes de STIRLING, si nous remplaçons encore une fois r par $r+1$,

$$(r+1)\psi_r(x) = \frac{1}{(r+2)!} + \sum_{s=0}^{s=r-1} \frac{(x-s)\psi_s(x)}{(r-s+1)!}. \tag{13}$$

Une autre formule de ce genre peut être déduite de l'identité (2), en y posant $y = -1$, puis remplaçant x par $x+1$, ce qui donnera, en vertu de l'équation fonctionnelle (8) de l'article IV,

$$\left\{ \begin{aligned} &(x-r+3)\Psi_{2r}(x) = \\ &= \sum_{s=1}^{s=r} (-1)^{s-1} \binom{x-r+s+1}{s+1} (x+1)^s \Psi_{2r-2s}(x), \end{aligned} \right. \tag{14}$$

d'où, en introduisant les polynomes de STIRLING, puis posant $r+1$ au lieu de r,

$$\left\{ \begin{aligned} &(x-r+2)\psi_r(x) = \frac{(-1)^{r-1}(x+1)^r}{(r+2)!} \\ &+ \sum_{s=1}^{s=r} \frac{(-1)^{s-1}(x+1)^s(x-r+s)}{(s+1)} \psi_{r-s}(x), \end{aligned} \right. \tag{15}$$

formule qui est plus compliquée que la précédente.

XIII.—Applications des polynomes symétriques.

Remarquons que le polynome du degré n

$$f(x) = \left(x+\frac{1}{n+1}\right)\left(x+\frac{2}{n+1}\right)\ldots\left(x+\frac{n}{n+1}\right),$$

ou, ce qui est évidemment la même chose,

$$f(x) = x^n+\frac{C^1_{n+1}x^{n-1}}{n+1}+\ldots+\frac{C^r_{n+1}x^{n-r}}{(n+1)^r}+\ldots+\frac{C^n_{n+1}}{(n+1)^n}$$

est symétrique, savoir qu'il satisfait à l'équation fonctionnelle

$$f(-x-1) = (-1)^n f(x),$$

les formules générales, valables pour les polynomes symétriques [1], donnent immédiatement, dans le cas spécial dont il s'agit ici,

$$(1)\quad \left(1-(-1)^r\right)C^r_{n+1} = \sum_{s=0}^{s=r-1}(-1)^s\binom{n-s}{r-s}(n+1)^{r-s}\,C^s_{n+1}$$

$$(2)\quad C^{2r+1}_{n+1} = \sum_{s=0}^{s=2r}(-1)^s\binom{n-s}{2r-s+1}\left(\frac{n+1}{2}\right)^{2r-s+1}C^s_{n+1}$$

$$(3)\quad \begin{cases} (-1)^r\left(C^{2r+1}_{n+1}-\dfrac{(n+1)(n-2r)}{2}C^{2r}_{n+1}\right) = \\ = \displaystyle\sum_{s=0}^{s=r-1}(-1)^s\binom{n-2s+1}{2r-2s}(n+1)^{2r-2s}\,B_{r-s}\,C^{2s+1}_{n+1} \end{cases}$$

$$(4)\quad \begin{cases} (-1)^r\,C^{2r+1}_{n+1} = \\ = \displaystyle\sum_{s=0}^{s=r}(-1)^s\binom{n-2s}{2r-2s+1}\left(\frac{n+1}{2}\right)^{2r-2s+1}T_{r-s+1}\,C^{2s}_{n+1}, \end{cases}$$

où les B_s et les T_s désignent les nombres de Bernoulli respectivement les coefficients des tangentes.

On sait que ces quatre formules sont équivalentes, de sorte qu'une seule de ces formules entraîne nécessairement

[1] Voir mon Traité élémentaire des Nombres de Bernoulli, articles XXI et XXIV.

les trois autres. De plus, il est évident que les formules en question permettent de calculer les G_{n+1}^{2r+1} à l'aide des G_{n+1}^{s} aux indices inférieurs.

Cela posé, la méthode ordinaire donnera immédiatement les identités algébriques

$$(5)\quad (1-(-1)^r)\Psi_{2r}(x) = \sum_{s=0}^{s=r-1}(-1)^s\binom{x-s}{r-s}(x+1)^{r-s}\Psi_{2s}(x)$$

$$(6)\quad \Psi_{4r+2}(x) = \sum_{s=0}^{s=2r}(-1)^s\binom{x-s}{2r-s-1}\left(\frac{x+1}{2}\right)^{2r-s-1}\Psi_{2s}(x)$$

$$(7)\quad \left\{\begin{array}{l}(-1)^r\left(\Psi_{4r+2}(x)-\dfrac{(x-1)(x-2r)}{2}\Psi_{4r}(x)\right) = \\ = \displaystyle\sum_{s=0}^{s=r-1}(-1)^s\binom{x-2s+1}{2r-2s}(x+1)^{2r-2s}B_{r-s}\Psi_{4s+2}(x)\end{array}\right.$$

$$(8)\quad \left\{\begin{array}{l}(-1)^r\Psi_{4r+2}(x) = \\ = \displaystyle\sum_{s=0}^{s=r}(-1)^s\binom{x-2s}{2r-2s+1}\left(\frac{x+1}{2}\right)^{2r-2s+1}T_{r-s+1}\Psi_{4s}(x),\end{array}\right.$$

où x désigne une variable complexe quelconque, ce qui donnera pour les polynomes de Stirling

$$(9)\quad \left(1-(-1)^r\right)\psi_r(x) = \frac{(x+1)^r}{(r+1)!} - \sum_{s=0}^{s=r-1}\frac{(-1)^s(x+1)^{r-s}}{(r-s)!}\psi_s(x)$$

$$(10)\quad \psi_{2r}(x) = \frac{(x+1)^{2r}}{(2r+1)!\,2^{2r-1}} - \sum_{s=0}^{s=2r-1}\frac{(-1)^s\psi_s(x)}{(2r-s)!}\left(\frac{x+1}{2}\right)^{2r-s}$$

$$(11)\quad \left\{\begin{array}{l}(-1)^r\left(\psi_{2r}(x)-\dfrac{x+1}{2}\psi_{2r-1}(x)\right) = \\ = \displaystyle\sum_{s=0}^{s=r-1}\frac{(-1)^s(x+1)^{2r-2s}}{(2r-2s)!}B_{r-s}\psi_{2s}(x)\end{array}\right.$$

$$(12)\quad \left\{ \begin{array}{l} (-1)^r \psi_{2r}(x) = \dfrac{T_{r+1}(x+1)^{2r}}{(2r+1)!\,2^{2r+1}} - \\ \quad - \displaystyle\sum_{s=0}^{s=r-1} \frac{(-1)^s T_{r-s}}{(2r-2s-1)!} \left(\frac{x+1}{2}\right)^{2r-2s-1} \psi_{2s+1}(x). \end{array} \right.$$

Il saute aux yeux que la dernière de ces quatre formules récursives est la plus simple, parce qu'elle permet de calculer $\psi_{2s}(x)$, seulement à l'aide des $\psi_{2r+1}(x)$ qui correspondent à des indices inférieurs. On voit du reste que les formules (9) et (10) sont identiques à (9) et (11) de l'article précédent.

Soit $x = 0$, il résulte, en vertu de (9), (10), (12),

$$(13)\qquad \frac{r-1}{2} = \sum_{s=1}^{<\frac{r-1}{2}} (-1)^{s-1} \binom{r+1}{2s} B_s$$

$$(14)\qquad 2r = \sum_{s=1}^{s=r} (-1)^{s-1} \binom{2r+1}{2s} 2^{2s} B_s$$

$$(15)\qquad T_{r+1} = \sum_{s=0}^{s=r-1} \binom{2r+1}{2s+1} 2^{2r-2s} B_{r-s} T_{s+1},$$

tandis que la formule (11) ne donne qu'une nouvelle détermination de la valeur numérique $\psi_{2r-1}(0)$.

La formule (13) n'est autre chose que la formule récursive de Jacques Bernoulli [1], faussement attribuée à Moivre [2] respectivement à Jacobi [3], selon que r est supposé pair ou impair; la formule (14) est due à Euler [4].

Revenons maintenant aux formules récursives contenant les polynomes $\Psi_{2m}(x)$, puis posons $x = -n-1$, où n

[1] Ars conjectandi, p. 97; Bâle 1713.
[2] Miscellanea analytica; Londres 1730.
[3] Journal de Crelle, t. 11, p. 265; 1834.
[4] Opuscula analytica, t. II, p. 264—265; Pétrograde 1785.

désigne un positif entier, il résulte, pour les nombres de Stirling de seconde espèce,

$$(16)\quad \left(1-(-1)^r\right)\mathfrak{S}^r_{n+1} = \sum_{s=0}^{s=r-1}(-1)^s\binom{n+s+1}{r-s}n^{r-s}\,\mathfrak{S}^s_{n+1}$$

$$(17)\quad \mathfrak{S}^{2r+1}_{n+1} = \sum_{s=0}^{s=2r}(-1)^s\binom{n+s+1}{2r\quad s+1}\binom{n}{2}^{2r-s+1}\mathfrak{S}^s_{n+1}$$

$$(18)\quad \begin{cases} (-1)^r\left(\mathfrak{S}^{2r+1}_{n+1} - \dfrac{n(n+2r+1)}{2}\mathfrak{S}^{2r}_{n+1}\right) - \\ \qquad = \displaystyle\sum_{s=0}^{s=r-1}(-1)^s\binom{n+2s}{2r-2s}n^{2r-2s}\,B_{r-s}\,\mathfrak{S}^{2s+1}_{n+1} \end{cases}$$

$$(19)\quad \begin{cases} (-1)^r\,\mathfrak{S}^{2r+1}_{n+1} = \\ = \displaystyle\sum_{s=0}^{s=r}(-1)^s\binom{n+2s+1}{2r-2s+1}\binom{n}{2}^{2r-2s+1}T_{r-s+1}\,\mathfrak{S}^{2s}_{n+1}, \end{cases}$$

formules qui sont analogues à celles indiquées pour les nombres de Stirling de première espèce, et que l'on peut déduire aussi, à l'aide de la théorie des polynomes symétriques.[1]

XIV.—D'autres formules récursives.

Une autre classe de formules récursives peut être établie à l'aide de la définition des nombres de Stirling de première espèce, savoir

$$x(x+1)(x+2)\ldots(x+p) = x^{p+1} + C^1_{p+1}x^p + \ldots + C^p_{p+1}x.$$

A cet effet, posons, dans cette formule, successivement

$$x = 1,\ 2,\ 3,\ \ldots,\ n,$$

[1] Voir mon Traité élémentaire des Nombres de Bernoulli, article LXXXII.

puis ajoutons toutes les équations ainsi obtenues, il résulte, en vertu d'une formule élémentaire bien connue,

$$\frac{n(n+1)\ldots(n+p+1)}{p+2} =$$
$$= S_{p+1}(n) + C_{p+1}^{1} S_p(n) + \ldots + C_{p+1}^{p} S_1(n),$$

équation numérique qui est applicable, pourvu que n soit un positif entier quelconque, et qui est par conséquent équivalente à l'identité algébrique

$$\frac{x(x+1)\ldots(x+p+1)}{p+2} =$$
$$= \Phi_{p+2}(x) + C_{p+1}^{1} \Phi_{p+1}(x) + \ldots + C_{p+1}^{p} \Phi_2(x),$$

où x désigne une variable complexe quelconque.

Cherchons maintenant, dans cette identité algébrique, le coefficient de la puissance x^{p-r+2}, il résulte, pourvu que $2 \leq r \leq p$,

$$(1)\quad \left\{ \begin{array}{l} \dfrac{C_{p+2}^{r}}{p+2} = \dfrac{C_{p+1}^{r}}{p-r+2} + \dfrac{C_{p+1}^{r-1}}{2} - \\ \qquad - \displaystyle\sum_{s=1}^{\leq \frac{r}{2}} \frac{(-1)^s B_s}{2s} \binom{p-r+2s+1}{2s-1} C_{p+1}^{r-2s}; \end{array} \right.$$

le cas particulier $r = 1$ conduira à une trivialité, tandis que l'hypothèse $r = p+1$ donnera la formule de Schlömilch

$$(2)\qquad \frac{p!\,p}{2p+4} = \sum_{s=1}^{\leq \frac{p+1}{2}} (-1)^{s-1} B_s C_{p+1}^{p-2s+1},$$

savoir la formule (3) de l'article IX.

Revenons maintenant à la formule générale (1), puis éliminons, en vertu de la formule récursive

$$C_{p+2}^{r} = C_{p+1}^{r} + (p+1) C_{p+1}^{r-1},$$

le nombre C_{p+2}^{r}, il résulte, après une simple réduction,

$$(3)\quad \left\{\begin{array}{l} \dfrac{r\,C_{p+1}^{r}}{p+2} = \dfrac{p\,(p-r-2)}{2p+4}\,C_{p-1}^{r-1} \\ \qquad + \displaystyle\sum_{s=1}^{\leq \frac{r}{2}} (-1)^{s} \binom{p-s+2s+1}{2s} B_s\, C_{p+1}^{r-2s}\,, \end{array}\right.$$

équation numérique qui est valable, pourvu que p soit un positif entier plus grand que $r-1$, et qui est par conséquent équivalente à l'identité algébrique

$$(4)\quad \left\{\begin{array}{l} \dfrac{r\,\Psi_{2r}(x)}{x+2} = \dfrac{x\,(x+r-2)}{2x+4}\,\Psi_{2r-2}(x) \\ \qquad + \displaystyle\sum_{s=1}^{\leq \frac{r}{2}} (-1)^{s} \binom{x-r+2s+1}{2s} B_s\,\Psi_{2r-4s}(x)\,, \end{array}\right.$$

où x désigne une variable complexe quelconque, la valeur $x=-2$ étant exclue, mais multiplions par $x+2$ les deux membres de la formule en question, il est évident que cette valeur de x est aussi applicable.

Quant aux polynomes de Stirling, il faut, dans (4), étudier séparément les deux cas qui correspondent à des valeurs paires et impaires de n. Appliquons ensuite les valeurs de $\psi_m(0)$, indiquées dans l'article XI, il résulte finalement

$$(5)\quad \left\{\begin{array}{l} (r+1)\,(x-r)\,\psi_r(x) = \dfrac{x\,(x-r-1)}{2}\,\psi_{r-1}(x) - (x-2)\,\psi_r(0) - \\ \qquad - (x+2)\displaystyle\sum_{s=1}^{\leq \frac{r}{2}} \frac{(-1)^s\,(x-r+2s)}{(2s)!}\,B_s\,\psi_{r-2s}(x)\,, \end{array}\right.$$

où il faut supposer naturellement $r \geq 2$.

Posons encore, dans (4), $x=-p-1$, où p désigne un positif entier plus grand que 1, il résulte, pour les nombres de Stirling de seconde espèce,

$$(6)\quad \begin{cases} \dfrac{r\,\mathfrak{G}_{p+1}^{r}}{p-1} = \dfrac{(p+1)(p-r+3)\,\mathfrak{G}_{p+1}^{r-1}}{2p-2} \\ \qquad -\sum_{s=1}^{\leq \frac{r}{2}} (-1)^s \binom{p+r-1}{2s} B_s\, \mathfrak{G}_{p+1}^{r-2s}. \end{cases}$$

formule récursive qui est analogue à la formule (3) concernant les nombres de STIRLING de première espèce.

Revenons maintenant à la définition transcendante des polynomes de STIRLING, la formule (22) de l'article VI donnera pour $y = 0$

$$(x+2)\,\psi_r(x+1) =$$

$$= (x+1)\,\psi_r(x) + \psi_r(0) + (x+1)\cdot \sum_{s=0}^{r-1} \psi_s(0)\,\psi_{r-s-1}(x),$$

d'où, en éliminant, en vertu de l'équation fonctionnelle (3) de l'article XI, la fonction $\psi_r(x+1)$,

$$(7)\quad \begin{cases} (r+1)\,\psi_r(x) = \\ = \dfrac{x+1}{2}\,\psi_{r-1}(x) - \psi_r(0) - (x-1)\cdot \sum_{s=1}^{\leq \frac{r}{2}} \dfrac{(-1)^s B_s}{(2s)!}\,\psi_{r-2s}(x), \end{cases}$$

formule récursive qui est analogue à (5), mais plus simple, parce que le facteur de $\psi_r(x)$ est une constante.

Or, on voit qu'une combinaison de ces deux formules conduira à une relation encore plus simple. En effet, multiplions par $x+2$ les deux membres de (7), puis soustrayons (5) de la formule ainsi obtenue, il résulte

$$(8)\quad \begin{cases} (r+1)(r+2)\,\psi_r(x) = \left(\left(\dfrac{r}{2}+1\right)x+1\right)\psi_{r-1}(x) + \\ \qquad -(x-2)\cdot \sum_{s=1}^{\leq \frac{r}{2}} \dfrac{(-1)^s (r-2s-1)\,B_s}{(2s)!}\,\psi_{r-2s}(x). \end{cases}$$

Il est évident que cette dernière formule récursive con-

duira à des relations numériques intéressantes. Soit, en premier lieu, r un nombre impair, savoir $r = 2n-1$, l'hypothèse $x = 0$ donnera

$$(9) \qquad n(2n+1)B_n = \sum_{s=1}^{s=n-1} \binom{2n}{2s} (2n-2s) B_s B_{n-s},$$

où, ce qui est évidemment la même chose,

$$(10) \qquad \left(n+\frac{1}{2}\right) B_n = \sum_{s=1}^{s=n-1} \binom{2n-1}{2s} B_s B_{n-s}.$$

Écrivons maintenant dans l'ordre inverse les termes qui figurent au second membre de cette dernière formule, nous aurons

$$(10\text{ bis}) \qquad \left(n+\frac{1}{2}\right) B_n = \sum_{s=1}^{s=n-1} \binom{2n-1}{2s-1} B_s B_{n-s},$$

d'où, en ajoutant (10) et (10 bis),

$$(11) \qquad (2n+1) B_n = \sum_{s=1}^{s=n-1} \binom{2n}{2s} B_s B_{n-s};$$

cette formule est due à Euler[1], et il est évident qu'elle est équivalente à (10) et à (10 bis).

En second lieu, posons $r = 2n$, puis différentions par rapport à x les deux membres de la formule ainsi obtenue, nous aurons, en posant $x = 0$,

$$(2n+1)(2n+2)\psi'_{2n}(0) = (n+1)\psi_{2n-1}(0) + \psi'_{2n-1}(0) +$$
$$+ \frac{(-1)^n B_n}{(2n)!\,2} - 2 \cdot \sum_{s=1}^{s=n-1} \frac{(-1)^s (2n-2s+1) B_s}{(2s)!} \psi'_{2n-2s}(0),$$

d'où, en multipliant par $(2n)!$, puis appliquant les valeurs numériques indiquées dans l'article XI,

[1] Opuscula analytica, t. II, p. 266; Pétrograde 1785.

$$\frac{(2n+1)(n+1)(2n-1)\,B_n}{2n} = \frac{2n+1}{2}\,B_n + (-1)^{n-1}(2n)!\,\psi'_{2n-1}(0) + \\ + \sum_{s=1}^{s=n-1}\binom{2n}{2s}\left(2n-2s-\frac{1}{2n-2s}\right)B_s\,B_{n-s}.$$

Appliquons ensuite la formule (9), il résulte finalement

$$(12)\quad (-1)^n(2n)!\,\psi'_{2n-1}(0) = \frac{2n+1}{2n}\,B_n - \sum_{s=1}^{s=n-1}\binom{2n}{2s}\frac{B_s\,B_{n-s}}{2s}.$$

expression qui est formellement différente de celle que nous avons développée dans la formule (13 bis) de l'article XI. Or, ajoutons les deux formules en question, il résulte

$$(13)\quad \begin{cases} \left(\dfrac{1}{1}+\dfrac{1}{2}+\dfrac{1}{3}+\ldots+\dfrac{1}{2n}\right)B_n = \\ \qquad = \displaystyle\sum_{s=1}^{s=n-1}\left(\binom{2n}{2s}-1\right)\frac{B_s\,B_{n-s}}{2s}; \end{cases}$$

cette formule récursive pour les nombres de Bernoulli, très curieuse ce me semble, est certainement nouvelle.

Nous avons encore à étudier le cas particulier de la formule (23) de l'article VI, qui correspond à $y = 0$, savoir

$$(14)\quad \begin{cases} B_{n+1}(x) - (x+1)\cdot\displaystyle\sum_{s=0}^{s=n}(-1)^s\,\psi_s(x)\,B_{n-s}(x) = \\ = (x-1)\psi_n(x) - (-1)^n\psi_n(0) - (x+1)\cdot\displaystyle\sum_{s=0}^{s=n-1}(-1)^s\,\psi_s(0)\,\psi_{n-s-1}(x); \end{cases}$$

appliquons ensuite la formule (7), il résulte pour n pair

$$(15)\quad \begin{cases} B_{2n+1}(x) = (2x-2n+1)\,\psi_{2n}(x) + \\ \quad + (x+1)\cdot\displaystyle\sum_{s=0}^{s=2n-1}(-1)^s\,\psi_s(x)\,B_{2n-s}(x), \end{cases}$$

tandis que l'hypothèse n impair donnera de même

$$(15\text{ bis})\quad \left\{ \begin{array}{l} B_{2n}(x) \qquad\qquad 2n\,\psi_{2n-1}(x) \\ +(x+1)\cdot\displaystyle\sum_{s=0}^{s=2n-2}(-1)^s\,\psi_s(x)\,B_{2n-s-1}(x). \end{array} \right.$$

Dans ces deux formules il faut supposer $n \geq 1$.

Différentions maintenant par rapport à x les deux membres de (15 bis), puis posons $x = -1$, nous aurons

$$2n\,\psi'_{2n-1}(-1) = -\sum_{s=0}^{s=n-2}\psi_{2s+1}(-1)\,B_{2n-2s-2}(-1),$$

d'où, après une simple réduction,

$$(-1)^n\,2n\cdot(2n)!\,\psi'_{2n-1}(-1) = \sum_{s=1}^{s=n-1}\binom{2n}{2s}\frac{B_s\,B_{n-s}}{2s},$$

ou, ce qui est la même chose,

$$(-1)^n(2n)!\,\psi'_{2n-1}(-1) = \sum_{s=1}^{s=n-1}\binom{2n-1}{2s}\frac{B_s\,B_{n-s}}{2s(2n-2s)},$$

savoir

$$(-1)^n(2n)!\,\psi'_{2n-1}(-1) = \sum_{s=1}^{s=n-1}\binom{2n-1}{2s-1}\frac{B_s\,B_{n-s}}{2s(2n-2s)},$$

ce qui donnera finalement le résultat numérique curieux

$$(16)\quad (-1)^n(2n)!\,\psi'_{2n-1}(-1) = \frac{1}{2}\cdot\sum_{s=1}^{s=n-1}\binom{2n}{2s}\frac{B_s\,B_{n-s}}{2s(2n-2s)},$$

analogue à (12) du reste. Ajoutons ces deux formules, nous aurons

$$(17)\quad \psi'_{2n-1}(0) + 2n\,\psi'_{2n-1}(-1) = \frac{(-1)^n(2n-1)}{(2n)!\,2n}B_n,$$

ce qui donnera, pour les coefficients $\alpha_{2n-1,\,r}$ du polynome $\psi_{2n-1}(x)$,

$$(18)\quad \alpha_{2n-1,\,2n-2} + \sum_{s=0}^{s=2n-3}\frac{(-1)^s(2n-s-1)}{2n+1}\alpha_{2n-1,\,s} = \frac{(-1)^n\,B_n}{(2n)!\,2n}.$$

d'où, en vertu de la formule (5) de l'article XI

$$(19)\qquad \sum_{s=0}^{s=2n-3}(-1)^s(s+2)\,a_{2n-1,\,s}=\frac{(-1)^{n-1}(n+1)(2n-1)}{(2n)!\,n}B_n.$$

Ces résultats numériques montrent clairement que les coefficients compliqués $a_{n,r}$ sont intimement liés aux nombres de Bernoulli.

XV.— Expressions explicites.

La littérature extrêmement riche qui traite plus ou moins profondément les nombres de Stirling, présente plusieurs expressions explicites des nombres susdits.

Cauchy[1] a par exemple pris pour point de départ la formule (18) de l'article VI; introduisant, au premier membre, la série de puissances de la fonction exponentielle, puis appliquant la formule polynomiale, il obtient une expression explicite de $\mathfrak{S}_{p-1}^{n}$, mais cette expression est parfaitement inapplicable pour un calcul pratique. Remarquons que la même méthode donnera, en vertu de la formule (17) de l'article susdit, une expression explicite de C_{p-n}^{n}.

Schläfli[2] a donné d'autres expressions explicites des C_n^p — but this law is a very complicated one — dit Cayley[3] et avec raison, parce que Schläfli exprime les C_n^p par des sommes multiples. Or, les formules de Cayley ne disent rien sur la nature des nombres C_n^p, et c'est la même chose pour les expressions, formées à l'aide des déterminants, données par von Zeipel.[4] C'est pourquoi nous avons à établir

[1] Exercises de Mathématiques, IIIe année, p. 145—147; Paris 1828.
[2] Journal de Crelle, t. 43, p. 1—22; 1852.
[3] Quarterly Journal of Mathematics, t. 3, p. 308; 1860.
[4] Annuaire de l'Université de Lund (en suédois) 1870.

d'autres expressions explicites des polynomes $\Psi_{2n}(x)$ et des $\psi_n(x)$.

A cet effet, prenons pour point de départ l'identité [1]

$$(1) \qquad f(x) = x\binom{x+n}{n} \cdot \sum_{s=0}^{s=n} \frac{(-1)^s}{x+s}\binom{n}{s} f(-s),$$

où $f(x)$ désigne un polynome du $n^{\text{ième}}$ degré par rapport à x.

Étudions tout d'abord la fonction $\Psi_{2n}(x)$, les valeurs numériques

$$\Psi_{2n}(0) = 0, \quad \Psi_{2n}(-1) = 0, \quad \Psi_{2n}(-s) = \mathfrak{S}_s^n, \qquad s > 2,$$

donnent immédiatement, en vertu de (1),

$$(2) \qquad \Psi_{2n}(x) = x\binom{x+2n}{2n} \cdot \sum_{s=2}^{s=2n} \frac{(-1)^s}{x+s}\binom{2n}{s} \mathfrak{S}_s^n.$$

On voit que cette formule est un peu compliquée, parce qu'elle n'indique pas directement que $\Psi_{2n}(x)$ est divisible par le produit

$$(x-1)(x-2)\ldots(x-n+1);$$

c'est-à-dire que nous aurons les formules numériques

$$(3) \qquad \sum_{s=2}^{s=2n} \frac{(-1)^s}{r+s}\binom{2n}{s} \mathfrak{S}_s^n = 0, \qquad 1 \leq r \leq n-1,$$

tandis que la valeur

$$\Psi_{2n}(n) = C_{n+1}^n = n!$$

donnera de même, en vertu de (2),

$$(3 \text{ bis}) \qquad \frac{(n-1)!\, n!\, (2n)!}{(3n)!} = \sum_{s=2}^{s=2n} \frac{(-1)^s}{n+s}\binom{2n}{s} \mathfrak{S}_s^n.$$

Il saute aux yeux que la formule (2) ne donne aucune expression simple des polynomes $\psi_n(x)$. Or, appliquons

[1] Voir mon Traité élémentaire des Nombres de Bernoulli, article VIII.

l'identité (1), les valeurs numériques

$$\psi_{2n}(0) = 0, \quad \psi_{2n}(-1) = 0,$$

$$\psi_n(-s) = \frac{(-1)^{n-1}(s-2)!\,\mathfrak{S}_s^{n+1}}{(s+n)!}, \qquad s \geq 2,$$

donnent immédiatement

$$(4) \quad \psi_{2n}(x) = x\binom{x-2n}{2n}\sum_{s=2}^{s=2n}\frac{(-1)^s}{x+s}\binom{2n}{s}\frac{(s-2)!\,\mathfrak{S}_s^{2n+1}}{(2n+s)!}.$$

Quant à $\psi_{2n-1}(x)$, il résulte de même, en vertu des valeurs numériques

$$\psi_{2n-1}(0) = \frac{(-1)^{n-1}B_n}{(2n)!}, \quad \psi_{2n-1}(-1) = \frac{(-1)^n B_n}{(2n)!\,2n},$$

indiquées dans l'article XI, l'expression explicite

$$(4\text{ bis}) \begin{cases} \psi_{2n-1}(x) = x\binom{x+2n-1}{2n-1}\left(\dfrac{(-1)^{n-1}B_n\big((2n+1)x+2n\big)}{(2n)!\,2n\,x(x+1)} - \right. \\ \left. -\displaystyle\sum_{s=2}^{s=2n-1}\frac{(-1)^s}{x+s}\binom{2n-1}{s}\frac{(s-2)!\,\mathfrak{S}_s^{2n}}{(2n+s-1)!}\right), \end{cases}$$

expression qui est plus compliquée que (4). Dans la formule (4 bis) il faut naturellement supposer $n > 2$; soit $n = 1$, la somme qui figure au second membre est à supprimer.

Appliquons maintenant les valeurs de $\psi'_{2n}(0)$ et de $\psi'_{2n}(-1)$, indiquées dans l'article XI, il résulte, en vertu de (4), les deux expressions explicites des nombres de BERNOULLI:

$$(5) \quad \sum_{s=2}^{s=2n}\frac{(-1)^s}{(s-1)s^2}\binom{4n}{2n-s}\mathfrak{S}_s^{2n+1} = \frac{(-1)^{n-1}(2n-1)}{4n}\binom{4n}{2n}B_n$$

$$(5\text{ bis}) \quad \sum_{s=2}^{s=2n}\frac{(-1)^s}{(s-1)^2 s}\binom{4n}{2n-s}\mathfrak{S}_s^{2n+1} = \frac{(-1)^{n-1}}{2}\binom{4n}{2n}B_n,$$

d'où, en soustrayant,

$$(6)\qquad \sum_{s=\frac{2}{2}}^{s=2n} \frac{(-1)^s}{(s-1)^2 s^2} \binom{4n}{2n-s} \mathfrak{S}_s^{2n+1} = \frac{(-1)^{n-1}}{2n} \binom{4n-1}{2n} B_n .$$

Quant aux polynomes $\Psi_{2m}(x)$, la formule

$$\Psi_{2m}(x) = (x+1)\, x\, (x-1) \ldots (x-m+1)\, \psi_{m-1}(x)$$

donnera immédiatement, en vertu des formules (4), des expressions explicites, dont celle de $\Psi_{4n}(x)$ devient un peu compliquée.

Revenons maintenant à la formule (4) de l'article VI, savoir

$$\left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-x-1} = 1 + \sum_{r=1}^{r=\infty} \binom{x+r}{r} \left(1 - \frac{1-e^{-\alpha}}{\alpha}\right)^r,$$

nous avons démontré, dans l'article susdit, qu'il existe, pour une valeur fixe mais quelconque de x, un nombre positif k, tel que la série qui figure au second membre est uniformément convergente, pourvu que $|\alpha| \leq k$, de sorte qu'il est permis d'ordonner formellement, d'après des puissances ascendantes, de α, la série en question.

A cet effet, appliquons l'identité

$$\left(1 - \frac{1-e^{-\alpha}}{\alpha}\right)^r = \sum_{s=0}^{s=r} (-1)^s \binom{r}{s} \left(\frac{1-e^{-\alpha}}{\alpha}\right)^s,$$

il est évident que le coefficient $A_{r,n}$ de la puissance α^n, provenant de ce terme, deviendra, en vertu de la définition (1) de l'article VI,

$$(7)\qquad A_{r,n} = \sum_{s=1}^{s=r} (-1)^{s-1} \binom{r}{s} s\, \psi_{n-1}(-s-1),$$

ou, ce qui est la même chose,

$$(7\text{ bis})\quad A_{r,n} = \frac{r!\,(n-1)!}{(n+r)!} \cdot \sum_{s=1}^{s=r} \frac{(-1)^{n+s}}{s!} \binom{n+r}{r-s} \mathfrak{S}_{s+1}^{n},$$

ce qui donnera l'expression explicite de $\psi_{n-1}(x)$:

$$(8)\qquad (x+1)\,\psi_{n-1}(x) = \sum_{s=n}^{s=n} \binom{x+s}{s} A_{s,n}, \qquad n \geq 1.$$

Cela posé, cherchons, au second membre de (8), le coefficient α_r de la fonction $\psi_{n-1}(-r-1)$, il résulte, en vertu de (7),

$$\alpha_r = (-1)^{r-1} r \cdot \sum_{s=0}^{s=n-r} \binom{r+s}{r} \binom{x+r+s}{r+s},$$

ou, ce qui est la même chose,

$$\alpha_r = (-1)^{r-1} r \binom{x+r}{r} \cdot \sum_{s=0}^{s=n-r} \binom{x+r+s}{s},$$

de sorte que nous aurons

$$\alpha_r = (-1)^{r-1} r \binom{x+r}{r} \binom{x-n+1}{n-r},$$

savoir

$$\alpha_r = \frac{(-1)^{r-1} r\,(x+1)}{x+r+1} \binom{x+n+1}{n} \binom{n}{r},$$

ce qui donnera

$$(9)\qquad \left\{ \begin{array}{c} \psi_{n-1}(x) = \\ = \binom{x-n+1}{n} \cdot \sum_{s=0}^{s=n-1} \frac{(-1)^s (s+1)}{x-s-2} \binom{n}{s} \psi_{n-1}(-s-2). \end{array} \right.$$

Appliquons ensuite les formules

$$\psi_{n-1}(n+s-1) = \frac{(-1)^{n-1}(n-s-1)!\,\mathfrak{S}^n_{n-s-1}}{(2n-s)!},$$

il résulte finalement, en vertu de (9),

$$(10)\qquad \psi_{n-1}(x) = \frac{\omega_n(x+2)}{(2n)!} \cdot \sum_{s=0}^{s=n-1} (-1)^s \binom{2n}{s} \frac{\mathfrak{S}^n_{n-s-1}}{x-n-s+1},$$

d'où, en vertu des valeurs de $\psi_{2n-1}(0)$ et $\psi_{2n-1}(-1)$, les expressions explicites de nombres de Bernoulli:

$$(11)\qquad \frac{(-1)^{n-1}}{2n+1}\binom{4n}{2n}B_n = \sum_{s=0}^{s=2n-1}(-1)^s\binom{4n}{s}\frac{\mathfrak{S}_{2n}^{2n-s+1}}{2n-s-1}$$

$$(11\text{ bis})\qquad \frac{(-1)^n}{2n}\binom{4n}{2n}B_n = \sum_{s=0}^{s=2n-1}(-1)^s\binom{4n}{s}\frac{\mathfrak{S}_{2n}^{2n-s+1}}{2n-s},$$

tandis que les valeurs numériques

$$\psi_{2n}(0) = \psi_{2n}(-1) = 0$$

donnent de même

$$(12)\qquad 0 = \sum_{s=0}^{s=2n-2}(-1)^s\binom{4n-2}{s}\frac{\mathfrak{S}_{2n-s}^{2n-1}}{2n-s}$$

$$(12\text{ bis})\qquad 0 = \sum_{s=0}^{s=2n-2}(-1)^s\binom{4n-2}{s}\frac{\mathfrak{S}_{2n-s}^{2n-1}}{2n-s-1}.$$

Revenons maintenant à la formule générale (10), une simple réduction donnera

$$(13)\qquad \left\{\begin{array}{c}\Psi_{2n}(x) = \\ = (x-n-1)\binom{x+n-1}{2n}\cdot\sum_{s=0}^{s=n-1}(-1)^s\binom{2n}{s}\frac{\mathfrak{S}_{n-s+1}^{n}}{x-n-s+1},\end{array}\right.$$

d'où, en supposant x égal au positif entier $p \geq n$,

$$(14)\qquad \left\{\begin{array}{c}C_{p+1}^{n} = \\ = (p-n+1)\binom{p+n+1}{2n}\cdot\sum_{s=0}^{s=n-1}(-1)^s\binom{2n}{s}\frac{\mathfrak{S}_{n-s+1}^{n}}{p-n-s+1};\end{array}\right.$$

cette formule, due à Schlömilch [1], représente évidemment l'expression explicite la plus simple connue pour les nombres de Stirling de première espèce.

[1] Journal de Crelle, t. 44, p. 352, 1852. Compendium der höheren Analysis, t. II, p. 28, 3e édition, Brunswick 1879.

XVI.— Séries de polynomes $\psi_n(x)$.

Remarquons que la série de puissances

$$\left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-x-1} = 1+(x+1)\cdot\sum_{n=0}^{n=\infty}\psi_n(x)\,\alpha^{n+1}$$

a son rayon de convergence égal à 2π, il résulte, quel que soit x,

$$\text{(1)}\qquad \limsup_{n=\infty}\left|\sqrt[n]{\psi_n(x)}\right| = \frac{1}{2\pi}.$$

Or, il est possible d'obtenir des résultats beaucoup plus généraux, concernant la valeur de $\psi_n(x)$, pour n extrêmement grand. En effet, soit C_r la circonférence, prise dans le sens direct, du cercle $|x| = r$, où $r < 2\pi$, l'intégrale de CAUCHY donnera

$$\left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-x-1} = \frac{1}{2\pi i}\cdot\int_{C_r}\frac{(1-e^{-t})^{-x-1}\,dt}{(t-\alpha)\,t^{-x-1}}.$$

Supposons maintenant $\Re(x) < 0$, l'intégrale curviligne susdite a un sens pour $r = 2\pi$ et représente par conséquent la fonction analytique dont il s'agit, ce qui donnera

$$(x+1)\,\psi_n(x) = \frac{1}{2\pi i}\cdot\int_{C_{2\pi}}\frac{(1-e^{-t})^{-x-1}\,dt}{t^{n-x}},$$

d'où, en posant $t = 2\pi e^{\theta i}$,

$$\text{(2)}\quad (2\pi)^{n-x}(x+1)\,\psi_n(x) = \int_0^{2\pi}\left(1-e^{-2\pi e^{\theta i}}\right)^{-x-1} e^{(x-n+1)\theta i}\,d\theta,$$

de sorte que nous aurons

$$\text{(3)}\quad (2\pi)^n\cdot|(x+1)\,\psi_n(x)| < \int_0^{2\pi}\left|\left(1-e^{-2\pi e^{\theta i}}\right)^{-x-1}\right| e^{|x|\theta}\,d\theta,$$

parce que nous avons supposé $\Re(x) < 0$.

Étudions maintenant l'ensemble $\mathfrak{E}$ des valeurs de x qui satisfont aux deux conditions

(4) $$\Re(x) \leqq -\delta < 0, \quad |x| \leqq \rho,$$

où δ et ρ sont des nombres positifs, dont le premier est arbitrairement petit, le second arbitrairement grand, l'intégrale définie qui figure au second membre de (3) a, dans l'ensemble $\mathfrak{E}$, une valeur maximum M, ce qui donnera, quel que soit l'indice n,

(5) $$|(x+1)\psi_n(x)| < \frac{M}{(2\pi)^n},$$

où nous aurons évidemment, pour δ très petit, $M > 1$.

Cela posé, prenons pour point de départ la formule récursive

$$(n+1)\psi_n(x) =$$

$$= \frac{x+1}{2}\psi_{n-1}(x) - \psi_n(0) + (x+1)\cdot\sum_{\mu=1}^{<\frac{n}{2}} \frac{(-1)^\mu B_\mu}{(2\mu)!}\psi_{n-2\mu}(x),$$

savoir la formule (7) de l'article XIV, puis appliquons la formule eulérienne[1] bien connue

$$\frac{B_\mu (2\pi)^{2\mu}}{(2\mu)!} = 2\left(\frac{1}{1^{2\mu}} + \frac{1}{2^{2\mu}} + \frac{1}{3^{2\mu}} + \cdots\right) = 2s_{2\mu},$$

nous aurons immédiatement, en vertu de (5),

(6) $$(n+1)(2\pi)^n|\psi_n(x)| < \pi M + (2\pi)^n|\psi_n(0)| + 2M\cdot\sum_{\mu=1}^{<\frac{n}{2}} s_{2\mu}.$$

Soit maintenant n un nombre pair, on trouvera

$$\psi_{2n}(0) = 0$$

$$s_{2\mu} \leqq s_2 = \frac{\pi^2}{6} < \pi,$$

ce qui donnera, en vertu de (6),

$$(2\pi)^{2n}|\psi_{2n}(x)| < \pi M.$$

[1] Introductio in analysin infinitorum, t. 1, § 168; 1748.

Soit, au contraire, n un nombre impair, nous aurons

$$(2\pi)^{2n+1}|\psi_{2n+1}(0)| = \frac{s_{2n+2}}{2\pi} \leq \frac{s_2}{2\pi} = \frac{\pi}{12} < \pi,$$

de sorte que nous trouvons, dans ce cas [1],

$$(2\pi)^{2n+1}|\psi_{2n+1}(x)| < \pi M,$$

de sorte que nous avons démontré la proposition suivante:

I. Supposons que la variable x appartienne à l'ensemble $\mathfrak{E}$, le polynome $\psi_n(x)$ satisfait, quel que soit l'indice n, à l'inégalité

$$(7) \qquad (2\pi)^n \cdot |\psi_n(x)| < \pi M,$$

où M désigne la valeur maximum de l'intégrale définie qui figure au second membre de (3).

Soit plus généralement $\mathfrak{E}_r$ l'ensemble des valeurs de x, déterminées par les deux conditions

$$(8) \qquad \Re(x) \leqq -r-\delta, \quad |x| \leqq \rho,$$

où r désigne un positif entier, nous aurons évidemment

$$(9) \qquad |x+r| < |x+r-1| < \ldots < |x+1| < |x|$$

$$(9\text{ bis}) \qquad |n-x| > |n-x+1| > \ldots > |n-x+r| > n.$$

Cela posé, nous avons à démontrer la proposition générale:

II. Supposons que la variable x appartienne à l'ensemble $\mathfrak{E}_r$, les polynomes de Stirling satisfont, quel que soit l'indice n, à l'inégalité

$$(10) \qquad (2\pi)^n |\psi_n(x)| < \frac{(2\pi+1)^r \rho^{r-1} \pi M}{n^r}.$$

[1] Nous supposons $M > 1$, ce qui est évident; or, désignant par M le plus grand des nombres 1 et la valeur maximum de l'intégrale qui figure au second membre de (3), on voit que les résultats suivants sont exacts, parce que la valeur numérique qui figure au premier membre de (7) est plus petite qu'une constante fixe.

Étudions tout d'abord le cas $r = 1$, l'équation fonctionnelle

$$(x+2)\,\psi_n(x+1) = (x-n)\,\psi_n(x) + (x+1)\,\psi_{n-1}(x)$$

donnera immédiatement, pourvu que $x \neq n$,

$$(11) \quad |\psi_n(x)| < \frac{|(x+2)\,\psi_n(x+1)| + |(x+1)\,\psi_{n-1}(x)|}{|n-x|}.$$

Or, remarquons que les deux nombres x et $x+1$ appartiennent tous deux à l'ensemble $\mathfrak{E}$, pourvu que x appartienne à $\mathfrak{E}_1$, il résulte, en vertu de (7),

$$(2\pi)^n \cdot |\psi_n(x)| < \frac{(2\pi+1)\pi M}{|n-x|},$$

ce qui donnera immédiatement, en vertu de (9 bis), la formule (10) pour $r = 1$.

Supposons maintenant valable, dans l'ensemble $\mathfrak{E}_m$, l'inégalité

$$(12) \qquad (2\pi)^n |\psi_n(x)| < \frac{(2\pi+1)^m \rho^{m-1} \pi M}{|n-m+1-x|^m},$$

puis supposons que x appartienne à l'ensemble $\mathfrak{E}_{m+1}$, il est évident que $x+1$ appartient toujours à l'ensemble $\mathfrak{E}_m$, de sorte que la formule est applicable si nous y remplaçons x par $x+1$, ce qui donnera, en vertu de (11),

$$(2\pi)^n |\psi_n(x)| <$$

$$< \frac{(2\pi+1)^m \rho^m \pi M}{|n-x|} \left(\frac{1}{|n-m-x|^m} + \frac{2\pi}{|n-m+1-x|^m} \right),$$

de sorte que nous aurons, en vertu de (9 bis),

$$(2\pi)^n |\psi_n(x)| < \frac{(2\pi+1)^{m+1} \rho^m \pi M}{|n-m-x|^{m+1}} < \frac{(2\pi+1)^{m+1} \rho^m \pi M}{n^{m+1}}.$$

Quant aux valeurs positives de $\mathfrak{R}(x)$, désignons par E_r l'ensemble des valeurs de x qui satisfont aux deux conditions

$$r-1-\delta \leqq \Re(x) \leqq r-\delta, \quad |x| \leqq \rho, \tag{13}$$

où r désigne un positif entier, nous aurons cette autre proposition, supplémentaire à la précédente:

III. Supposons que la variable x appartienne à l'ensemble E_r, les polynomes de Stirling satisfont, quel que soit n, à l'inégalité

$$(2\pi)^n |\psi_n(x)| < (2\pi+1)^r n^r \pi M. \tag{14}$$

Étudions tout d'abord le cas, où $x+1$ appartient à l'ensemble E_1, il est évident que x doit appartenir à l'ensemble $\mathfrak{E}$, savoir satisfaire aux conditions

$$-1-\delta \leqq \Re(x) \leqq -\delta < 0, \quad |x| \leqq \rho,$$

de sorte que l'équation fonctionnelle des $\psi_n(x)$ donnera immédiatement

$$\begin{cases} (2\pi)^n |\psi_n(x+1)| < \\ < \left|\dfrac{n-x}{x+2}\right| (2\pi)^n |\psi_n(x)| + \dfrac{(2\pi)^n}{|x+2|} |(x+1)\psi_{n-1}(x)|, \end{cases} \tag{15}$$

et nous aurons par conséquent

$$|x+2| > |x+1|, \quad |x+2| > \rho-1$$
$$|n-x| \leqq n+\rho < (\rho-1)n.$$

ce qui donnera

$$(2\pi)^n |\psi_n(x+1)| < n\pi M + 2\pi^2 M < n(2\pi+1)\pi M,$$

savoir la formule (14), pour $r=1$.

Supposons maintenant vraie la formule générale (14), puis supposons que $x+1$ appartienne à l'ensemble E_{r+1}, de sorte que x est évidemment situé dans l'ensemble E_r, nous aurons par conséquent, en vertu de (15),

$$(2\pi)^n |\psi_n(x+1)| < (2\pi+1)^{r+1} n^{r+1} \pi M,$$

et la conclusion de r à $r+1$ est établie.

Combinons maintenant les deux propositions II et III, il résulte le théorème général:

IV. Supposons $|x| \leq \rho$, les polynomes de Stirling satisfont, quel que soit l'indice n, à l'inégalité

$$(16) \qquad (2\pi)^n |\psi_n(x)| < (2\pi+1)^\rho n^\rho \rho^{\rho-1} \pi M.$$

Cela posé, il est très facile de démontrer cette autre théorème:

V. Soit

$$(17) \qquad \limsup_{n=\infty} \left| \sqrt[n]{a_n} \right| = \sigma < 2\pi,$$

la série de polynomes de Stirling

$$(18) \quad F(x) = a_0\psi_0(x) + a_1\psi_1(x) + \ldots + a_n\psi_n(x) + \ldots$$

est absolument convergente, quel que soit x, et uniformément convergente, pourvu que $|x| \leq K$, de sorte que $F(x)$ est une transcendante entière.

La convergence absolue de la série (18) est une conséquence immédiate de (1) et (17), tandis que la convergence uniforme résulte de l'inégalité (16).

Quant à la série (18), elle ne peut jamais converger, pourvu que $\sigma > 2\pi$, tandis que l'hypothèse $\sigma = 2\pi$, exige des recherches ultérieures.

Posons par exemple, dans la formule

$$\left(\frac{1-e^{-\alpha}}{\alpha}\right)^{-x-1} = 1 + (x+1) \cdot \sum_{n=0}^{n=\infty} \psi_n(x)\, \alpha^{n+1}$$

$\alpha = 2\pi i$, la série ainsi obtenue est absolument et uniformément convergente, pourvu que x appartienne à l'ensemble $\mathfrak{E}_2$, de sorte que nous aurons

$$(19) \qquad \sum_{n=0}^{n=\infty} (-1)^n (2\pi)^{2n} \psi_{2n}(x) = 0$$

$$(19\text{ bis}) \qquad \sum_{n=0}^{n=\infty} (-1)^n (2\pi)^{2n+2} \psi_{2n+1}(x) = \frac{1}{x+1},$$

formules qui sont valables, pourvu que $\Re(x) < -2$.

Supposons maintenant remplie la condition (17), puis posons, conformément au développement (18),

$$(20) \qquad f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots,$$

la formule (2) donnera immédiatement

$$\frac{x+1}{(2\pi)^x} \cdot a_n \psi_n(x) = \int_0^{2\pi} \left(1 - e^{2\pi e^{\theta i}}\right)^{-x-1} e^{(x-1)\theta i} a_n \left(\frac{e^{-\theta i}}{2\pi}\right)^n d\theta,$$

de sorte que les deux fonctions analytiques $F(x)$ et $f(x)$ sont liées par l'équation intégrale

$$(21) \qquad \frac{x+1}{(2\pi)^x} \cdot F(x) = \int_0^{2\pi} \left(1 - e^{2\pi e^{\theta i}}\right)^{-x-1} e^{(x+1)\theta i} f\left(\frac{e^{-\theta i}}{2\pi}\right) d\theta,$$

où il faut supposer naturellement $\Re(x) < 0$.

On voit que cette formule intégrale n'est pas applicable pour les deux séries (19) et (19 bis), séries qui correspondent à l'hypothèse $\sigma = 2\pi$.

Citons encore, en terminant cet article, une intégrale curviligne qui représente le nombre $\mathfrak{S}_{n+1}^r$.

A cet effet, appliquons les valeurs limites

$$(22) \qquad \lim_{t=p} \frac{n!\, t^{n+r-1}}{(t-1)(t-2)\ldots(t-n)} = (-1)^{n-p} \binom{n}{p} p^{n+r},$$

puis désignons par C la circonférence d'un cercle ayant son centre dans l'Origine et son rayon plus grand que n, il résulte immédiatement

$$(23) \qquad \frac{1}{2\pi i} \cdot \int_C \frac{n!\, t^{n+r-1}\, dt}{(t-1)(t-2)\ldots(t-n)} = \mathfrak{S}_{n+1}^r.$$

Du reste, on voit que les valeurs limites (22) donnent cette autre formule plus générale

$$(24)\quad \frac{1}{2\pi i}\int_C \frac{n!\,f(t+\alpha)\,dt}{t(t-1)\dots(t-n)} = \sum_{s=0}^{s=n} (-1)^s \binom{n}{s} f(\alpha+n-s),$$

où $f(\alpha+t)$ est holomorphe sur la circonférence et dans l'intérieur du cercle C.

QUATRIÈME PARTIE

Résultats numériques.

XVII. — Les coefficients des $\psi_n(x)$.

Quant aux coefficients $\alpha_{n,r}$ du polynome $\psi_n(x)$, savoir

$$(1)\quad \psi_n(x) = \alpha_{n,0}\,x^n + \alpha_{n,1}\,x^{n-1} + \ldots + \alpha_{n,n-1}\,x + \alpha_{n,n},$$

il est évident que les expressions explicites, développées dans l'article XV, donnent des expressions explicites de ces coefficients.

Prenons par exemple pour point de départ la formule (8), savoir

$$(x+1)\,\psi_{n-1}(x) = \sum_{s=1}^{s=n} \binom{x+s}{s} A_{s,n},$$

où

$$A_{s,n} = \frac{r!\,(n-1)!}{(n+r)!} \cdot \sum_{r=1}^{r=s} \frac{(-1)^{n-s}}{s!} \binom{n+r}{r-s} \mathfrak{S}^n_{s+1},$$

puis posons pour abréger

$$(x+2)(x+3)\ldots(x+r) = c_{r,0}\,x^{r-1} + c_{r,1}\,x^{r-2} + \ldots + c_{r,r-1},$$

il résulte l'expression explicite

$$(2)\qquad \alpha_{n-1,p} = \sum_{s=0}^{s=p} \frac{c_{n-s,\,p-s}}{(n-s)!}\, A_{n-s,n}$$

qui est un peu compliquée.

La formule (10) de l'article XV, savoir

$$\psi_{n-1}(x) = \frac{\omega_n(x+2)}{(2n)!} \cdot \sum_{s=0}^{s=n-1} (-1)^s \binom{2n}{s} \frac{\mathfrak{S}^n_{n-s+1}}{x+n-s+1},$$

donnera un résultat plus simple.

En effet, posons pour abréger

$$\frac{\omega_n(x+2)}{x+k} = d_{n,0}^{(k)}\, x^{n-1} + d_{n,1}^{(k)}\, x^{n-2} + \ldots + d_{n,n-1}^{(k)},$$

où il faut supposer naturellement $2 \leq k \leq n+1$, nous aurons

$$(3) \qquad \alpha_{n-1,p} = \frac{1}{(2n)!} \cdot \sum_{s=0}^{s=n-1} (-1)^s \binom{2n}{s} \mathfrak{S}_{n,s+1}^{n} d_{n,p}^{(n-s-1)},$$

expression qui est aussi assez compliquée pour un calcul direct des coefficients $\alpha_{n-1,p}$. De plus, on voit que les expressions explicites que nous venons d'indiquer contiennent des nombres de Stirling qu'il s'agit précisément de déterminer.

Cela posé, nous avons à revenir à des formules récursives qui permettent de déterminer successivement les coefficients $\alpha_{n,p}$.

A cet effet, remarquons que nous avons déjà donné une formule récursive de ce genre, savoir la formule (6 bis) de l'article V, formule qui se présente aussi sous la forme

$$(4) \qquad \left| \begin{array}{c} (2r-p+2)\,\alpha_{r,p} = \\ = \alpha_{r-1,p} + \alpha_{r-1,p-1} - \sum_{s=0}^{r=p-1} \binom{r-s}{p-s} \frac{r+p-2s-2}{p-s+1} \alpha_{r,s}, \end{array} \right.$$

et qui n'est autre chose qu'une équation aux différences finies, linéaire et du premier ordre, pourvu que les coefficients

$$\alpha_{r,0},\ \alpha_{r,1},\ \ldots,\ \alpha_{r,p-1}$$

soient connus pour des valeurs quelconques de r.

Or, il est évident que les formules récursives que nous venons de développer, pour les $\mathfrak{b}_n(x)$, donnent plusieurs autres équations du même genre.

En premier lieu, appliquons la formule (7) de l'article XIV, savoir

$$(r+1)\,\psi_r(x) =$$

$$= \frac{x+1}{2}\,\psi_{r-1}(x) - \psi_r(0) + (x+1)\cdot\sum_{s=1}^{\leqq\frac{r}{2}}\frac{(-1)^s B_s}{(2s)!}\,\psi_{r-2s}(x),$$

nous aurons

$$(5)\quad \left\{\begin{array}{l} (r+1)\,\alpha_{r,p} = \\ = \dfrac{\alpha_{r-1,p}+\alpha_{r-1,p-1}}{2} + \displaystyle\sum_{s=1}^{\leqq\frac{p+1}{2}}\frac{(-1)^s B_s}{(2s)!}\cdot\alpha_{r-2s,\,p-2s+1} + \\ \qquad + \displaystyle\sum_{s=1}^{\leqq\frac{p}{2}}\frac{(-1)^s B_s}{(2s)!}\cdot\alpha_{r-2s,\,p-2s}, \end{array}\right.$$

où il faut, pour $p = 1$, supprimer la dernière somme qui figure au second membre. Multiplions maintenant par 2 les deux membres de (5), puis soustrayons (4), il résulte la formule récursive

$$(6)\quad \left\{\begin{array}{l} p\,\alpha_{r,p} = \displaystyle\sum_{s=0}^{s=p-1}\binom{r-s}{p-s}\frac{r+p-2s+2}{p-s+1}\,\alpha_{r,s} - \\ \qquad 2\cdot\displaystyle\sum_{s=1}^{\leqq\frac{p+1}{2}}\frac{(-1)^s B_s}{(2s)!}\,\alpha_{r-2s,\,p-2s+1} + \\ \qquad + 2\cdot\displaystyle\sum_{s=1}^{\leqq\frac{p}{2}}\frac{(-1)^s B_s}{(2s)!}\,\alpha_{r-2s,\,p-2s}, \end{array}\right.$$

formule qui permet de déterminer directement le coefficient $\alpha_{r,p}$, pourvu que les coefficients précédents soient connus.

On aura par exemple

$$(7)\qquad \alpha_{r,1} = -\frac{r(r-5)}{(r+1)!\,2^{r+1}\cdot 6}.$$

En second lieu, étudions la formule (13) de l'article XII, savoir

$$(r+1)\,\psi_r(x) = \frac{1}{(r+2)!} + \sum_{s=0}^{s=r-1} \frac{(x-r+s-1)\,\psi_{r-s-1}(x)}{(s+2)!},$$

nous aurons de même

$$(8)\quad \left\{ \begin{array}{l} (r+1)\,\alpha_{r,p} = \\ = \displaystyle\sum_{s=0}^{s=p} \frac{\alpha_{r-s-1,\,p-s}}{(s+2)!} - \sum_{s=0}^{s=p-1} \frac{(r-s-1)\,\alpha_{r-s-1,\,p-s-1}}{(s+2)!} \end{array} \right.,$$

ce qui donnera, en vertu de (4),

$$(9)\quad \left\{ \begin{array}{l} p\,\alpha_{r,p} = -\alpha_{r-1,\,p-1} + \\ \quad + \displaystyle\sum_{s=0}^{r=p-1} \binom{r-s}{p-s} \frac{r+p-2s+2}{p-s+1}\,\alpha_{r-s-1,\,p-s-1} - \\ \quad + \displaystyle\sum_{s=1}^{s=p} \frac{2\,\alpha_{r-s-1,\,p-s}}{(s+2)!} - \sum_{s=0}^{s=p+1} \frac{(2r-2s-2)\,\alpha_{r-s-1,\,p-r-1}}{(s+2)!} \end{array} \right.,$$

En dernier lieu, prenons pour point de départ la formule (7) de l'article II, savoir

$$n\,C_n^r = \sum_{s=0}^{s=r} (-1)^s \binom{n-r+s}{s+1} C_{n+1}^{r-s},$$

ou, ce qui est la même chose,

$$x\,\Psi_{2r}(x-1) = \sum_{s=0}^{s=r} (-1)^s \binom{x-r+s}{s+1} \Psi_{2r-2s}(x);$$

introduisons ensuite les $\psi_m(x)$, puis posons $r+1$ au lieu de r, nous aurons

$$(10)\quad x\,\psi_r(x-1) = (x+1) \cdot \sum_{s=0}^{s=r} \frac{(-1)^s\,\psi_{r-s}(x)}{(s+1)!} - \frac{(-1)^r}{(r+2)!},$$

ce qui donnera

$$(11)\quad \left\{ \begin{array}{l} \displaystyle\sum_{s=0}^{s=p} (-1)^{p-s+1} \binom{r-s}{p-s+1} \alpha_{r,s} = \sum_{s=0}^{s=p} \frac{(-1)^s\,\alpha_{r-s,\,p-s}}{(s+1)!} - \\ \qquad + \displaystyle\sum_{s=1}^{s=p+1} \frac{(-1)^s\,\alpha_{r-s,\,p-s+1}}{(s+1)!} \end{array} \right.,$$

de sorte que nous aurons, en vertu de (8),

$$(12)\quad \left\{\begin{aligned}(p+1)\alpha_{r,p} &= \sum_{s=0}^{s=p-1}(-1)^s\binom{r-p+s+1}{s+2}\alpha_{r,p-s-1}+\\ &+\sum_{s=0}^{s=p-1}\frac{2\,\alpha_{r-2s-2,\,p-2s-1}}{(2s+3)!}-\\ &-\sum_{s=0}^{s=p-1}\frac{(r-s+1+(-1)^s)\,\alpha_{r-s-1,\,p-s-1}}{(s+2)!},\end{aligned}\right.$$

où il faut supposer naturellement $p \geq 1$.

XVIII.— Nature analytique des coefficients.

Les formules récursives que nous venons d'établir sont très compliquées pour un calcul successif des coefficients $\alpha_{r,p}$. Or, les formules en question donnent des éclaircissements intéressants concernant la nature analytique des $\alpha_{r,p}$.

A cet effet, posons, conformément à l'expression obtenue pour $\alpha_{r,1}$,

$$(1)\qquad \alpha_{r,p} = \frac{\beta_{r,p}}{(r+1)!\,2^{r+1}},$$

nous aurons évidemment

$$(2)\qquad \beta_{r,1} = -\frac{r(r-5)}{6}, \qquad \beta_{r,0} = 1,$$

et la conclusion ordinaire de p à $p+1$ donnera, en vertu de la formule (12) de l'article précédent, la proposition générale:

I. Les nombres $\beta_{r,p}$, définis par la formule (1), se présentent toujours sous la forme d'un polynome entier du dégré $2p$ par rapport à r.

Posons, conformément à cette propriété de $\beta_{r,p}$,

$$(3)\qquad \alpha_{r,p} = \frac{\beta_{2p}(r)}{(r+1)!\,2^{r+1}},$$

il résulte, en vertu de la formule (4) de l'article précédent, que le polynome $\beta_{2p}(x)$ satisfait à l'équation aux différences finies du premier ordre

$$(4)\quad \begin{cases} (2x-p+2)\,\beta_{2p}(x) - (2x+2)\,\beta_{2p}(x-1) = \\ = (2x+2)\,\beta_{2p-2}(x-1) \sum_{s=0}^{s=p-1} \binom{x-s}{p-s} \frac{x+p-2s+2}{p-s+1}\,\beta_{2s}(x). \end{cases}$$

tandis que la formule (6) de l'article susdit donnera la formule récursive

$$(5)\quad \begin{cases} p\,\beta_{2p}(x) = \sum_{s=0}^{s=p-1} \binom{x-s}{p-s} \frac{x+p-2s+2}{p-s+1}\,\beta_{2s}(x) + \\ + \sum_{s=1}^{\leq \frac{p+1}{2}} (-1)^s\, 2^{2s+1} \binom{x+1}{2s}\, B_s\, \beta_{2p-4s+2}(x-2s+1) + \\ + (x+1) \sum_{s=1}^{< \frac{p}{2}} (-1)^s\, 2^{2s+1} \binom{x}{2s}\, B_s\, \beta_{2p-4s}(x-2s). \end{cases}$$

Remarquons maintenant que le coefficient $\alpha_{s,p}$ n'a aucun sens pour $0 \leq r \leq p-1$, il est évident que le polynome $\beta_{2p}(x)$ s'évanouira pour de telles valeurs de x, de sorte que nous aurons la proposition suivante:

II. Soit $p \geq 1$, le polynome $\beta_{2p}(x)$ est toujours divisible par le produit

$$x(x-1)\ldots(x-p+1),$$

de sorte que nous aurons

$$(6)\quad \beta_{2p}(x) = x(x-1)\ldots(x-p+1)\,\sigma_p(x), \quad \beta_0(x) = \sigma_0(x),$$

où $\sigma_p(x)$ est un polynome du degré p par rapport à (x).

On voit que cette propriété de $\beta_{1p}(x)$ ne se présente pas comme une conséquence immédiate des formules (4)

et (5). En effet, posons, dans (4), $x = 0$, nous aurons, en vertu de (6),

$$\beta_{2p}(-1) = -\beta_{2p-2}(-1),$$

ce qui donnera

$$\beta_{2p}(-1) = (-1)^p, \tag{7}$$

de sorte que nous aurons

$$\sigma_p(-1) = \frac{1}{p!}. \tag{8}$$

Supposons maintenant dès à présent donnée la valeur

$$\beta_{2p}(0) = 0, \quad p > 1,$$

la conclusion de p à $p+1$, donnera immédiatement, en vertu de (4), l'expression générale (6).

Quant aux polynomes $\sigma_p(x)$, définis par la formule (6), la valeur numérique

$$\alpha_{2p,\,2p} = 0, \quad p \geq 1,$$

donnera la proposition:

III. Soit $p \geq 1$, le polynome $\sigma_{2p}(x)$ est toujours divisible par $x - 2p$.

Introduisons maintenant, dans (4), les expressions tirées de (6), il résulte l'équation aux différences finies du premier ordre

$$\left\{\begin{array}{l} (2x-p+2)\,\sigma_p(x) = \\ -\dfrac{2(x+1)(x-p)}{x}\,\sigma_p(x-1) + \dfrac{2(x+1)}{x}\,\sigma_{p-1}(x-1) - \\ -\displaystyle\sum_{s=0}^{s=p-1} \frac{x+p-2s+2}{(p-s-1)!}\,\sigma_s(x), \end{array}\right. \tag{9}$$

qui, suppléée par la valeur numérique (8), détermine parfaitement le polynome $\sigma_p(x)$, pourvu que les polynomes précédents soient connus.

La formule (5) donnera de même

$$(10)\quad \left\{\begin{aligned} p\,\sigma_p(x) &= \sum_{s=0}^{p-1} \frac{x+p-2s+2}{(p-s+1)!}\cdot \sigma_s(x) + \\ &+ (x+1)\cdot \sum_{s=1}^{\frac{p+1}{2}} \frac{(-1)^s\, 2^{2s+1}\, B_s}{(2s)!}\cdot \sigma_{p-2s+1}(x-s+1) + \\ &+ (x+1)\cdot \sum_{s=1}^{\frac{p}{2}} \frac{(-1)^s\, 2^{2s+1}\, B_s}{(2s)!}\cdot \sigma_{p-2s}(x-2s). \end{aligned}\right.$$

Nous renonçons aux autres formules récursives pour les $\sigma_p(x)$, obtenues des formules (9) et (12) de l'article précédent, parce qu'elles deviennent aussi compliquées que celles que nous venons de développer.

Revenons encore une fois à la formule (6) qui nous donne lieu à une remarque sur nos recherches précédentes. En effet, dans notre étude des nombres de Stirling du rang $n+1$ et de l'ordre p, savoir les C_{n+1}^{p} et les $\mathfrak{S}_{n+1}^{p}$, nous avons remplacé le positif entier n par une variable continue x, ce qui à donné les polynomes $\Psi_{2p}(x)$. En introduisant ensuite les polynomes de Stirling, savoir

$$\Psi_{2p}(x) = (x+1)\,x\,(x-1)\ldots(x-p+1)\,\psi_{p-1}(x),$$

la détermination des $\Psi_{2p}(x)$ à l'aide des $\psi_{p-1}(x)$ renferme une factorielle du rang p, dont l'étude est précisément le problème que nous nous sommes proposé de résoudre.

De plus, l'étude directe du polynome $\psi_r(x)$, savoir l'étude de ces coefficients $\alpha_{r,p}$ nous conduira de nouveau, comme le montre clairement la formule (6), à une factorielle du rang p.

Ces remarques montrent clairement qu'une étude approfondie des nombres de Stirling est un problème extrêmement compliqué.

XIX.—Les premiers polynomes $\psi_n(x)$.

En appliquant la formule (9) de l'article précédent j'ai trouvé les cinq premiers des fonctions $\sigma_p(x)$, savoir

$$\sigma_0(x) = 1$$

$$\sigma_1(x) = -\frac{x-5}{6}$$

$$\sigma_2(x) = \frac{(x-2)(x-11)}{3^2 \cdot 8}$$

$$\sigma_3(x) = -\frac{1}{6^4 \cdot 5}(5x^3 - 120x^2 + 619x - 336)$$

$$\sigma_4(x) = \frac{x-4}{6!\,6^3}(5x^3 - 170x^2 + 1271x + 150);$$

comme contrôle on aura la formule (8) de l'article précédent, savoir

$$\sigma_p(-1) = \frac{1}{p!}.$$

Quant aux coefficients des polynomes de Stirling, savoir

$$\psi_r(x) = \alpha_{r,0}x^r + \alpha_{r,1}x^{r-1} + \ldots + \alpha_{r,r-1}x + \alpha_{r,r},$$

on aura

$$\alpha_{r,p} = \frac{\sigma_p(r)}{(r-p)!\,(r+1)\,2^{r+1}},$$

de sorte que les résultats précédents nous permettent de calculer successivement les cinq premiers des $\psi_r(x)$.

Or, le calcul des polynomes susdits peut être étendu plus loin à l'aide des formules récursives que nous venons d'établir, et parmi lesquelles la formule (7) de l'article XIV est la plus commode.

En appliquant les valeurs des premiers nombres de Bernoulli, savoir

$$B_1 = \frac{1}{6},\quad B_2 = \frac{1}{30},\quad B_3 = \frac{1}{42},\quad B_4 = \frac{1}{30},\quad B_5 = \frac{5}{66},$$

j'ai calculé les sept premiers des polynomes de STIRLING, savoir

$$\psi_0(x) = \frac{1}{2}$$

$$\psi_1(x) = \frac{1}{4!}(3x+2)$$

$$\psi_2(x) = \frac{x(x+1)}{4!\,2}$$

$$\psi_3(x) = \frac{1}{6!\,2^3}(15x^3+15x^2-10x-8)$$

$$\psi_4(x) = \frac{x(x+1)}{6!\,2^4}(3x^2-x-6)$$

$$\psi_5(x) = \frac{1}{9!\,2^3}(63x^5-315x^3-224x^2+140x+96)$$

$$\psi_6(x) = \frac{x(x+1)}{9!\,2^4}(9x^4-18x^3-57x^2+34x+80),$$

ou, ce qui est la même chose,

$$\psi_0(x) = \frac{1}{2}$$

$$\psi_1(x) = \frac{x}{8}+\frac{1}{12}$$

$$\psi_2(x) = \frac{x^2}{48}+\frac{x}{48}$$

$$\psi_3(x) = \frac{x^3}{384}+\frac{x^2}{384}-\frac{x}{576}-\frac{1}{720}$$

$$\psi_4(x) = \frac{x^4}{3840}+\frac{x^3}{5760}-\frac{7x^2}{11520}-\frac{x}{1920}$$

$$\psi_5(x) = \frac{x^5}{46080}-\frac{x^3}{9216}-\frac{x^2}{12960}+\frac{x}{20736}+\frac{1}{30240}$$

$$\psi_6(x) = \frac{x^6}{645120}-\frac{x^5}{645120}-\frac{5x^4}{387072}-\frac{23x^3}{5806080}+$$
$$+\frac{19x^2}{967680}+\frac{x}{72576}.$$

On voit que $\psi_5(x)$ ne contient pas la puissance x^4, ce qui est une conséquence de la valeur de $\sigma_1(x)$, savoir $\sigma_1(5) = 0$.

XX.—Tables de Stirling et de M. Bertelsen.

Quant au calcul direct des nombres de Stirling, la méthode la plus commode est un calcul successif, à l'aide des formules récursives (15) de l'article I.

Abstraction faite des valeurs évidentes

$$C_n^1 = 1, \quad C_n^0 = 1, \quad n \geq 0,$$

on trouve pour les C_n^p :

p	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	p
1	3	6	10	15	21	28	36	1
2	2	11	35	85	175	322	546	2
3		6	50	225	735	1960	4536	3
4			24	274	1624	6769	22449	4
5				120	1764	13132	67284	5
6					720	13068	118124	6
7						5040	109584	7
8							40320	8

p	$n=10$	$n=11$	$n=12$	p
1	45	55	66	1
2	870	1320	1925	2
3	9450	18150	32670	3
4	63273	157773	357423	4
5	269325	902055	2637558	5
6	723680	3416930	13339535	6
7	1172700	8409500	45995730	7
8	1026576	12753576	105258076	8
9	362880	10628640	150917976	9
10		3628800	120543840	10
11			39916800	11

p	$n=13$	$n=14$	$n=15$	p
1	78	91	105	1
2	2717	3731	5005	2
3	55770	91091	143325	3
4	749463	1474473	2749747	4
5	6926634	16669653	37312275	5
6	44990231	135036473	368411615	6
7	206070150	790943153	2681453775	7
8	657206836	3336118786	14409322928	8
9	1414014888	9957703756	56663366760	9
10	1931559552	20313753096	159721605680	10
11	1486442880	26596717056	310989260400	11
12	479001600	19802759040	392156797824	12
13		6227020800	283465647360	13
14			87178291200	14

Quant aux nombres

$$\mathfrak{C}_{n+1}^{p} = \frac{1}{n!} \cdot \sum_{s=0}^{s=n} (-1)_s \binom{n}{s} (n-s)^{n+p},$$

nous aurons

$$\mathfrak{C}_{p}^{0} = 1, \quad \mathfrak{C}_{1}^{p} = 0, \quad p \geq 1$$

$$\mathfrak{C}_{2}^{p} = 1, \quad p \geq 0$$

$$\mathfrak{C}_{3}^{p} = 2^{p+1} - 1.$$

Abstraction faite de ces valeurs spéciales, on trouvera pour les $\mathfrak{C}_{n}^{p}$:

p	$n=3$	$n=4$	$n=5$	$n=6$	p
1	3	6	10	15	1
2	7	25	65	140	2
3	15	90	350	1050	3
4	31	301	1701	6951	4
5	63	966	7770	42525	5

p	$n=3$	$n=4$	$n=5$	$n=6$	p
6	127	3025	34105	246730	6
7	255	9330	145750	1379400	7
8	511	28501	611501	7508501	8
9	1023	86526	2532530	40075035	9
10	2047	261625	10391745	210766920	10
11	4095	788970	42355950	1096190550	11
12	8191	2375101	171798901	5652751651	12
13	16383	7141686	694337290	28958095545	13
14	32767	21457825	2798806985	147589284710	14

p	$n=7$	$n=8$	$n=9$	p
1	21	28	36	1
2	266	462	750	2
3	2646	5880	11880	3
4	22827	63987	159027	4
5	179487	627396	1899612	5
6	1323652	5715424	20912320	6
7	9321312	49329280	216627840	7
8	63436373	408741333	2141764053	8
9	420693273	3281882604	20415995028	9
10	2734926558	25708104786	189036065010	10
11	17505749898	197462483400	1709751003480	11
12	110687251039	1492924634839	15170932662679	12
13	693081601779	11143554045652	132511015347084	13
14	4306078895384	82310957214948	1142399079991620	14

p	$n=10$	$n=11$	p
1	45	55	1
2	1155	1705	2
3	22275	39325	3
4	359502	752752	4
5	5135130	12662650	5

p	$n = 10$	$n = 11$	p
6	67128490	193754990	6
7	820784250	2758334150	7
8	9528822303	37112163803	8
9	106175395755	477297033785	9
10	1144614626805	5917584964655	10
11	12011282644725	71187132291275	11
12	123272476465204	835143799377954	12
13	1241963303533920	9593401297313460	13
14	12320068811796900	108254081784931500	14

p	$n = 12$	$n = 13$	p
1	66	78	1
2	2431	3367	2
3	66066	106470	3
4	1479478	2757118	4
5	28936908	62022324	5
6	512060978	1256328866	6
7	8391004908	23466951300	7
8	129413217791	411016633391	8
9	1900842429486	6833042030178	9
10	26826851689001	108823356051137	10
11	366282500870286	1672162773483930	11
12	4864251308951100	24930204590758260	12
13	63100165695775560	362262620784874680	13
14	802355904438462660	5149507353856958820	14

p	$n = 14$	p
1	91	1
2	4550	2
3	165620	3
4	4910178	4
5	125854638	5

p	$n = 14$	p
6	2892439160	6
7	61068660380	7
8	1204909218331	8
9	22496861868481	9
10	401282560341390	10
11	6888836057922000	11
12	114485073343744260	12
13	1850568574253550060	13
14	29206898819153109600	14

Stirling [1] a calculé les nombres C_n^p de $n = 1$ à $n = 9$, il indique la valeur $C_9^6 = 105056$, ce qui est faux, parce que nous aurons $C_9^6 = 118124$, faute de calcul qui a été observée par Binet [2], dans son grand Mémoire sur les intégrales eulériennes. Stirling [3] a aussi calculé une petite table des nombres $\mathfrak{S}_n^p$, il indique faussement $\mathfrak{S}_8^2 = 461$, tandis que nous aurons $\mathfrak{S}_8^2 = 462$.

Les autres valeurs numériques, données dans les tables précédentes, sont calculées par M. N.-P. Bertelsen; une partie des tables de M. Bertelsen est publiée par feu M. Thiele [4] qui indique, qu'une copie des tables complètes est conservée dans la bibliothèque de l'Observatoire de Copenhague. Or, une telle copie n'étant pas trouvable dans la bibliothèque susdite, M. Bertelsen [5] m'a montré l'amabilité de faire, une seconde fois, les calculs nécessaires pour suppléer aux résultats publiés par Thiele, concernant les nombres $\mathfrak{S}_n^p$.

[1] Methodus differentialis, p. 11; Londres 1730.

[2] Journal de l'École Polytechnique, cahier 26, p. 231; 1839.

[3] Methodus differentialis, p. 8.

[4] Interpolationsrechnung, p. 31—32. Leipsic 1909.

[5] Je prie mon ami d'agréer mes vifs remerciements et pour ses nouvelles calculations et pour sa correction des épreuves de cet article.

Quant aux nombres

$$\mathfrak{A}_n^m = \sum_{s=0}^{s=n} (-1)^s \binom{n}{s} (n-s)^m,$$

savoir

$$\mathfrak{A}_n^m = 0, \quad n > m$$

$$\mathfrak{A}_n^m = n! \, \mathfrak{S}_{n+1}^{m-n}, \quad m > n,$$

ce qui donnera

$$\mathfrak{A}_1^m = 1, \quad m > 1, \quad \mathfrak{A}_n^n = n!,$$

J.-F.-W. Herschel [1] a calculé cette petite table des valeurs des $\mathfrak{A}_n^p$:

p	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	p
2	2	0	0	0	0	2
3	6	6	0	0	0	3
4	14	36	24	0	0	4
5	30	150	240	120	0	5
6	62	540	1560	1800	720	6
7	126	1806	8400	16800	15120	7
8	254	5796	40824	126000	191520	8
9	510	18150	186480	834120	1905120	9
10	1022	55980	818520	5103000	16435440	10

p	$n=7$	$n=8$	$n=9$	$n=10$	p
7	5040	0	0	0	7
8	141120	40320	0	0	8
9	2328480	1451520	362880	0	9
10	29635200	30240000	16329600	3628800	10

Plus tard Grunert [2] a calculé une partie de cette même table des $\mathfrak{A}_n^p$.

[1] A collection of examples of the applications of the calculus of finite differences, Cambridge 1820. La table de Herschel est reproduite dans les Nouvelles Annales des Mathématiques, t. 13, p. 272; 1854.

[2] Mathematische Abhandlungen, p. 71; Altona 1822.

TABLE DES MATIÈRES

Pages

Introduction 3

Première Partie. Les nombres de Stirling.

I. Définitions et propriétés fondamentales 11
II. Formules récursives générales 16
III. Applications élémentaires 19
IV. Introduction d'une variable continue 21
V. Les polynomes de Stirling 29

Deuxième Partie. De la fonction exponentielle.

VI. Nouvelles définitions des $\varphi_n(x)$ 33
VII. Sur une série d'Euler 40
VIII. Deux séries de Lagrange 43
IX. Applications diverses 48
X. D'autres applications 52

Troisième Partie. Les polynomes de Stirling.

XI. Évaluations des valeurs numériques 59
XII. Applications des nombres de Stirling 63
XIII. Applications des polynomes symétriques 68
XIV. D'autres formules récursives 71
XV. Expressions explicites 78
XVI. Séries de polynomes $\varphi_n(x)$ 84

Quatrième Partie. Résultats numériques.

XVII. Les coefficients des $\varphi_n(x)$ 92
XVIII. Nature analytique des coefficients 96
XIX. Les premiers des polynomes $\varphi_n(x)$ 100
XX. Tables de Stirling et de M. Bertelsen 102

Forelagt paa Modet d. 9. Maj 1919.
Færdig fra Trykkeriet d. 30. November 1920.

Q
62
D28
bd. 2

Danske videnskabernes selskab, Copenhagen
Matematisk-fysiske meddelelser

Physical & Applied Sci. Serials

PLEASE DO NOT REMOVE
CARDS OR SLIPS FROM THIS POCKET

UNIVERSITY OF TORONTO LIBRARY

STORAGE

#1